SUMMARY AND CONCLUSIONS

Data envelopment analysis is a linear programming tool implemented to measure efficiency scores, and to establish efficient targets to the decision making units. The approach is deterministic but not stochastic. In DEA it is customary to obtain efficiency scores in an attempt to project inefficient production plan (Input and output vector of inefficient DMU) on to the envelopment frontier, projections require choice of a distance function, radial or non-radial.

Choice of a distance function is not trivial in data envelopment analysis. Short run target setting requires radial distance functions as tools for projection. In short neither input mix nor output mix can be altered which implies that technique cannot be changed. Short run targets should be short, since efficiency gains are mostly possible through gains of managerial efficiency, possible through proportional reduction of inputs and augmentation of outputs. In DEA literature envelopment frontiers extensively used are convex frontiers, Charnes et al., 1978; Banker et al., 1984) and non-convex frontiers (Tulkens 1993; De Prins et al., 1984). For short run projections the most appropriate frontier is the non-convex frontier. Let $P_{CCR}, P_{BCC}$ and $P_{FDH}$ be respectively, the production possibility sets formulated by CCR, BCC and Free Disposable Hull (non-convex production possibility set). These sets are related as follows:

$$P_{FDH} \subseteq P_{BCC} \subseteq P_{CCR}$$

The radial targets based on the non-convex envelopment frontier are the shortest, thus, for very short run the appropriate distance function is radial and the envelopment frontier is non-convex.
The above graph depicts one input and one output production process. The straight line that emanates from the origin is Constant Returns to Scale (CCR) frontier. The line segments EA, AC and CD constitute the variable returns to scale frontier (BCC) and the stair case envelope determined by A, F, C, G and D is the Free Disposable Hull (FDH) non-convex frontier. Under CCR formulation A and B are efficient, whereas under BCC formulation E, A, C and D are efficient. But under FDH envelopment E, A, C, D, F and G are efficient. Since FDH envelopment is inner most, FDH projections (input/output oriented) are the shortest. We refer to these targets as ‘very short run’ targets. FDH frontier can be viewed as expost Production Frontier. To measure short run efficiency and to set short run targets for the inefficient decision making units, the most appropriate frontier is the BCC frontier that provides pure technical efficiency scores.

The present study refers to short run performance of Indian Public Sector Banks:

We assume that output expansion in short run is not possible, but proportional input reduction is possible. Under this hypothesis, the possibility of proportionate input reduction is examined for each of the 27 Indian Public Sector Banks. These banks control over 70 percent of Commercial Banks business. These banks are wide spread into rural India, and found to be more labour intensive than private and foreign sector banks. The size of the public sector banks measured in terms of total assets is much higher than private and foreign sector banks. Inspite of
the growing NPAs which may retard bank efficiency, the Public Sector Banks are found significantly more efficient than private and foreign sector banks (Kumar and Gulati, 2008; Sen Sarma 2005; Srivatsava 2006; Subramanyam and CS Reddy, 2008).

Due to the versatality in their activities their share in commercial bank business, we choose PSBs for efficiency performance study.

The BCC – DEA formulation, under input orientation may be presented as follows:

\[
\bar{\theta}_{BCC} = \min \theta - \varepsilon \left( \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right)
\]

such that

\[
\sum_{j=1}^{n} \lambda_j x_{ij} = \theta x_{ij} - s_i^-, \ i \in M
\]

\[
\sum_{j=1}^{n} \lambda_j y_{jr} = y_{jr} + s_r^+, \ r \in S
\]

\[
\sum_{j=1}^{n} \lambda_j = 1
\]

\[
\lambda_j \geq 0, \ j \in N
\]

where \(x_{ij}\): \(i^{th}\) input of \(j^{th}\) DMU

\(y_{jr}\): \(r^{th}\) output of \(j^{th}\) DMU

\(s_i^-\): \(i^{th}\) input slack

\(s_r^+\): \(r^{th}\) output slack

\(\varepsilon\): Non-Archimedean quantity

\(\lambda_j\): Intensity parameter of \(j^{th}\) DMU

\(j_0\) refers to \(DMU_{j_0}\) for which the optimization problem is being solved. \(DMU_{j_0}\) is said to be efficient if and only if,
\[ \bar{\theta}_{\text{BCC}} = 1 \]

This to happen all slacks must vanish at the optimum.

**Summary of BCC Scores**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.9572</td>
</tr>
<tr>
<td>SD</td>
<td>0.0405</td>
</tr>
<tr>
<td>Min</td>
<td>0.8856</td>
</tr>
<tr>
<td>Max</td>
<td>1.0000</td>
</tr>
<tr>
<td>Range</td>
<td>0.1144</td>
</tr>
</tbody>
</table>

The public sector banks experienced 4.3 percent of input losses on the average due to input pure technical inefficiency. The null hypothesis that the input and output plans of these banks arrive from efficient bank population is rejected with \( P < 0.01 \), (one percent level of significance).

**Distribution of BCC Efficiency Scores**

All the efficiency scores are found to distribute over the narrow interval (0.88, 1.00]. The frequency distribution of the BCC efficiency scores is more or less flat over the interval (0.88, 0.98), but efficiency concentration is in the interval [0.98, 1.00], 41 percent of the efficiency scores are found occurring in this interval.
The efficiency distribution is positively skewed

Canara Bank is the least scale efficient Public Sector Bank. Like the BCC efficiency score distribution, the input scale efficiency distribution is also found positively skewed. 81 percent public sector banks have their input scale efficiencies fallen in the interval [0.98, 1.0]. These are 22 out of 27 PSBs.

SECOND STAGE DATA ENVELOPMENT ANALYSIS

In second stage we try to explain interbank differences as measured by BCC pure technical efficiency scores, for which regression approach is pursued. Choice of regression model for second stage DEA is not trivial econometric problem. The BCC scores are free from scale effects, which lie in the interval [0, 1]. The standard linear regression, whose parameters can be estimated by the method of ordinary least squares, if applied to BCC efficiency scores, cannot prevent the predictions falling out of the fractional range.

differences reflected in CCR/BCC efficiency scores selecting appropriate environmental variables.

Net NPAs, ratio of Net NPAs to Net Advances, Off Balance Sheet Business earnings, size of the bank, number of bank branches, ownership viewed as a limited independent variable, were some of the environmental variables used to explain inter bank differences.

This study selected net NPAs, size of the bank measured by Total Assets and earnings by virtue of off balance sheet business, to explain BCC technical efficiency scores.

Fractional Regression equations introduced by Papke and Wooldridge (1996); Ramalho, Ramalho and Murtarea (2010) proposed generalized Fractional Regression equations; Logit and Probit regression models, Latent dependent variable regression, linear probability regression model, stochastic frontier regression model (Banker and Natarajan 2008), Bootstrap regression (Simar and Willson, 2007) are the alternative regression models implemented to explain inter bank differences reflected by input technical efficiency scores' variation.

Kumar and Gulati (2008) regressed environmental variables on binary dependent variable whose 0 and 1 values are generated by CCR overall technical efficiency scores. These scores belong to Indian Public Sector Banks.

For the second stage DEA we propose the following econometric regression model:

$$ y = \beta + \beta_N z_N + \beta_s z_s + \beta_o z_o + \varepsilon $$

where $z_N$: Net NPAs

$z_s$: Size as measured by Total Assets.

$z_o$: Earnings by virtue of off balance sheet business.

$\varepsilon$: Disturbance term.
y is binary dependent variable, which is defined on BCC pure technical scores as follows:

\[ y = 1, \text{ if } \theta_{\text{BCC}} = 1 \]
\[ = 0, \text{ if } \theta_{\text{BCC}} < 1 \]

This model can yield probit and logit probabilities based on alternative distributional assumptions regarding the disturbance term. If \( \varepsilon \) follows standard normal distribution, the binary regression model produces probits, on the other hand, if \( \varepsilon \) follows the logistic distribution then the binary regression model yields logistic probabilities.

The software implemented to estimate binary regression equation is SPSS. Regression fittings with all the three explanatory variables, pairs of them were unsatisfactory. One explanatory variable at a time was regressed on the binary regression variable. The regression of \( y \) on (total assets) size emerged to be a significant relation, with intercept significant at two percent level and the regression estimate \( y \) on size significant at 8 percent, the sign of the regression coefficient being found positive. For this model the number of correct predictions are about 78 percent. Maddala's, (1992) count \( R^2 \) therefore, is

\[ R^2_{\text{Court}} = 77.8 \]

This is a summary statistic. Another summary statistic is Nagalkerke \( R^2 \) whose value is 0.259 (NR=0.505).

-2 Log likelihood for this regression is, 28.798 and the predicted probabilities to be efficient are computed with the following formula:

\[ P(y = 1 / z_s) = \frac{e^{\hat{\beta}_s x_s}}{1 + e^{\hat{\beta}_s x_s}} = F(-2.005 + 0.000045 z_s) \]

The parameter estimate significance is based on Wald test.
For Indian Public Sector Banks we found a positive relationship between size of the bank and pure technical efficiency. State Bank of India and Punjab National Banks are the most robust banks among the nine BCC-efficient Banks. This inference is based on their predicted probabilities to remain efficient.

**RANKING OF EFFICIENT BANKS**

Petersen and Andersen (1993) introduced the concept of ‘Super Efficiency’ for the extremely efficient decision making units. The method is similar to Cook’s method in statistics to detect outliers. To find input super efficiency of an extremely efficient decision making unit, its input and output plan is removed from the reference technology and the modified frontier points are represented by,

\[
\left( \sum_{j \in I}^{n} \lambda_j x_j, \sum_{j \in J}^{n} \lambda_j y_j \right)
\]

The input and output plan of DMU$_{h}$ is projected onto the modified frontier.
For extremely efficient DMU, \( (x_{h_i}, y_{h_i}) \), represented by the extreme point B, the input set is given by \( L(\ y_{h_i} \) and the input insoquant is represented by the line segments AB, BC and CD. The input efficiency score of B is,

\[
\theta_{CCR} = \frac{OB}{OB} = 1
\]

If B \( (x_{h_i}, y_{h_i}) \) is removed from the reference set the input level set shrinks. The modified isoquant is given by the line segments AC and CD. A further radial projection of B onto the modified input frontier gives Super Efficiency for DMU,

\[
\theta_{Super}^{CCR} = \frac{OE}{OB} > 1
\]

Thus, we have, \( 0 < \theta_{CCR} \leq 1 \)

\[
\theta_{Super}^{CCR} > 1
\]

\( \theta_{Super}^{CCR} \) is a metric that indicates the stability of DMU, to remain efficient under input expansion. Larger Super efficiency implies greater superiority of the efficient DMU, among the extremely efficient DMU.
Andersen and Petersen (1993) introduced input super efficiency under CCR framework. This can be extended to BCC framework, while CCR-super efficiency problems are always feasible, the BCC-super efficiency problems are infeasible in certain cases (Seiford and Zhu, 1999) with the following complementarily being observed:

(i) If input SE-BCC problem is in feasible, then output SE-BCC problem is feasible.

(ii) If output SE-BCC problem is in feasible, then input SE-BCC problem is feasible.

Due to the presence of infeasibility for the purpose of ranking efficient DMUs, the BCC super efficiency approach is not recommended.

**GRAPHICAL ILLUSTRATION OF INFEASIBILITY**

![Infeasibility illustration](image)

**Input SE-BCC: Infeasibility (Seiford and Zhu, 1999)**

In the above figure input and output are measured along horizontal and vertical axes. At A input SE-BCC is feasible, but output SE-BCC is infeasible. At D output SE-BCC is feasible, but input SE-BCC is infeasible.

We use the predicted probabilities of efficient decision making units to remain efficient, in order to rank them, on the basis that larger is the probability to remain efficient greater is the superiority of the efficient bank.
Another alternative to rank the efficient decision making units is by means of solving the super efficiency problems based on directional distance function, that seeks input expansion and output contraction simultaneously.

\[
\beta_{j_0}^{Super} = \text{Max} \beta
\]

Such that

\[
\sum_{j \neq j_0}^{n} \lambda_j x_{ij} \leq (1-\beta) x_{ij}, \quad i \in M
\]

\[
\sum_{j \neq j_0}^{n} \lambda_j y_{rj} \geq (1+\beta) y_{rj}, \quad r \in S
\]

\[
\sum_{j \neq j_0}^{n} \lambda_j = 1
\]

\[
\lambda_j \geq 0, \quad j \neq j_0
\]

Cooper et al., (2007) show that this problem is always feasible

\[
\beta_{j_0}^{Super} < 0
\]

Smaller is \( \beta_{j_0}^{Super} \) greater is its ability to remain efficient under input expansion and output contraction. Two sets of ranks of the public sector banks are obtained. One set based on predicted probabilities to remain efficient and DDF based super efficiency values. To examine if the two sets agree, Spearman's rank correlation coefficient is computed.

\[
\hat{\rho} = 0.9499
\]

when tested against \( H_0: \rho = 1 \), the null hypothesis is accepted at \( P < 0.05 \).

**Efficient Input and Output Targets**

Data Envelopment Analysis not only provides efficiency scores, but also efficient input and output targets to inefficient decision making units. For providing such benchmarks, the additive model (Cooper et al., 1999) is solved for each Public Sector Commercial Bank.
Additive Model Under L₁-norm

\[
\text{Max } \left( \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{n} s_r^+ \right)
\]

such that

\[
\sum_{j=1}^{n} \lambda_j x_{ij} = x_{ij} - s_i^-, \quad i \in M
\]

\[
\sum_{j=1}^{n} \lambda_j y_{jr} = y_{jr} + s_r^+, \quad r \in S
\]

\[
\sum_{j=1}^{n} \lambda_j = 1
\]

\[
\lambda_j \geq 1
\]

For Bank \(_b\), the efficient targets are,

Input Targets: \( x_{ib}^* = x_{ib} - s_i^- \), \( i \in M \)

Output Targets: \( y_{rb}^* = y_{rb} + s_r^+ \), \( r \in S \)

RANGE ADJUSTED MEASURE

The optimal solution of additive problem estimates input and output slacks, consequently the efficient input and output targets, but fails to give a direct measure of efficiency such as, \( \theta_{CCR} \), \( \theta_{BCC} \), \( \theta_{DDF} \) or \( \theta_{FDH} \). However, the Range Adjusted Measure of efficiency (RAM) introduced by Cooper et al., (1999), can be computed for each commercial bank. Thus, RAM is a derived measure of efficiency.

\[ 0 \leq \text{RAM} \leq 1 \]

RAM efficiency scores are calculated to each Commercial Bank

**Summary Table for RAM**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.9982</td>
</tr>
<tr>
<td>SD</td>
<td>0.0018</td>
</tr>
<tr>
<td>Max</td>
<td>1.0000</td>
</tr>
<tr>
<td>Min</td>
<td>0.9933</td>
</tr>
</tbody>
</table>
**BRWZ Measure:** This efficiency measure was introduced by Brocket *et al.*, (1997). It is a derived measure that is units invariant, bound by the limits 0 and 1. This efficiency measure is evaluated for each Commercial Bank and the summary table is as follows:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>0.9098</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>0.0912</td>
</tr>
<tr>
<td><strong>Max</strong></td>
<td>1.0000</td>
</tr>
<tr>
<td><strong>Min</strong></td>
<td>0.6786</td>
</tr>
</tbody>
</table>

**Summary Table for BRWZ**

**Conclusions:**

1. The competition among the Public Sector Banks is fierce as evidenced by high CCR/BCC input technical efficiency scores, derived for their short run efficiency examination.

2. The PSB are highly efficient group of banks.

3. There are 7 efficient banks among 27 under CCR formulation.

4. There are 9 efficient banks among 27 under BCC formulation, confirmed by all the efficiency measurement approaches, namely, BCC input approach, Directional Distance function approach, RAM approach and BRWZ approach.

5. For all the above approaches efficient banks and their scores (unity) remain to be the same, but the scores of inefficient banks vary.

6. Since the inefficiency scores vary from one approach to another, for setting input and output targets the Additive Model under $L_1$- norm is the best one.

7. There is a positive relationship between pure technical efficiency scores and Bank size which is revealed by the fit of the Binary Regression.

8. The most robust is State Bank of India, followed by Punjab Notional Bank, among the seven (CCR) efficient banks.

9. The predicted probabilities for efficient decision making units to remain efficient are used for further discrimination of efficient Banks so that DEA discriminated power is enhanced.
10. Due to infeasibility input BCC super efficiency problems, BCC-SE approach is not useful for ranking of efficient decision making units.

Xue and Harker (2002) divide the super efficient DMUs into two categories, viz, super efficient and strongly super efficient. Strong Super Efficient DMUs have superiority over simply super efficient decision making units. However, this approach does not prevent infeasibility in input SE-BCC.

The directional Distance Function Approach yields Super Efficiency scores for each extremely efficient DMU, free from the problem of infeasibility with the help of this formulation the efficient DMUs are ranked.

11. To examine the degree of agreeability between the two sets of ranks, Spearman’s rank correlation is computed which is in conformity with the hypothesis that,

\[ H_0 : \rho = 1 \]

LIMITATIONS OF THE PRESENT STUDY

(i) For the second stage DEA to explain inter bank differences binary regression is implemented. In this approach the information contained in 27 efficiency scores is plugged into two number 0 and 1, there by the model suffers from loss of information. A better approach would be to implement the fractional regression equation. This approach does not sacrifice information.

(ii) The efficiency scores depend upon, how a commercial bank is modelled. Production model, intermediation model, model based on value added approach and model based on profit approach.

This study pursued profit approach to model the activities of a commercial bank.

(iii) Due to limitations of secondary data published in RBI bulletins we could not try to regress other environmental variables pursued in international studies of commercial banks, by different researchers.