



University of Connecticut

Department of Economics Working Paper Series

**Measures of Labor Use Efficiency from a Cost-Based Dual
Representation of the Technology: A Study of Indian Bank Branches**

by

Subhash C. Ray
University of Connecticut

Abhiman Das
Indian Institute of Management

Kankana Mukherjee
Babson College

Working Paper 2018-17
October 2018

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
<http://www.econ.uconn.edu/>

This working paper is indexed in RePEc, <http://repec.org>

Measures of Labor Use Efficiency from a Cost-Based Dual Representation of the Technology: A Study of Indian Bank Branches

Subhash C. Ray^{a*}

^aUniversity of Connecticut
Storrs CT 06269-1063, USA
subhash.ray@uconn.edu

Abhiman Das^b

^bIndian Institute of Management
Ahmedabad, India
abhiman@iima.ac.in

Kankana Mukherjee^c

^cBabson College
Babson Park, MA 02457, USA
kmukherjee@babson.edu

In this paper, we propose a representation of the production technology in the form of a *cost set* in the output and expenditure space as an alternative to the standard free disposal convex hull of input-output vectors. We show that when all units pay the same input prices, one can construct a free disposal convex hull of outputs and total expenditures to solve the cost minimization problem. We use the proposed model to evaluate the labor use efficiency of a sample of 325 branches of a major Indian public sector bank from four metropolitan cities across two years, 2008 and 2014. This is the first study in the Indian banking context to model the operations of branches using the *production approach*. Our empirical findings indicate that there is significant inefficiency in labor use in the branches and cost could be curtailed substantially by addressing overstaffing. Across the three types of labor, reducing the expenditure of clerks would have the highest impact for cost saving. We do find, however, that the extent of overspending on clerks has reduced in 2014, which apparently is a direct consequence of computerization of routine jobs. Efficiency varies across regions. In general, Chennai branches are more efficient than branches from other regions whereas Kolkata branches are the least efficient.

Keywords: Data envelopment analysis; Cost set; Group and industry frontiers; Indian bank branches; Labor use efficiency

JEL Classification: **G21, C61**

*Corresponding author. Tel: +1 860 486 3022; Fax: +1 860 486 4463.

Measures of Labor Use Efficiency from a Cost-Based Dual Representation of the Technology: A Study of Indian Bank Branches

1. Introduction

Since 1969, India has pursued one of the largest state-led bank credit programs ever attempted in a developing country. Starting with the nationalization of major commercial banks in 1969, the effort has been to create an extensive network of public sector banks and their branches in order to expand financial coverage throughout the country. The government has maintained the position that the role of the banking sector is to achieve the larger social and economic goals of the country and meet the financial needs of all the sections of society.¹

In the early 1990s, India instituted major economic reforms, a central aspect of which was financial sector reforms. The Reserve Bank of India (RBI) initiated measures to improve the efficiency and profitability of Indian banks. These measures included deregulation on entry of private and foreign banks to increase competition, allowing public sector banks to raise capital from the market, branch delicensing, deregulation on interest rates and other quantitative restrictions on bank portfolios, while at the same time introducing prudential regulations to improve the soundness of banks. Propelled by the economic reforms, the Indian economy grew rapidly, reaching a peak of over 9% annual average GDP growth rate during the years 2005-06 through 2007-08.

Another component of the banking reforms has been the gradual computerization of banking operations in India. Even as early as in 1982, C. Rangarajan, one of the Deputy Governors of the RBI was charged with the responsibility of introducing computers in Indian banking, starting with the RBI, on a modest scale. This, however, faced strong resistance from labor unions of public sector banks, especially in the Marxist state of West Bengal. In the immediate aftermath of the reforms, computers made a slow entry into Indian banking sector and although the newly licensed private sector banks were able to leverage technology in a big way, on the broader banking landscape computers had no significant presence. In fact the RBI started reporting the numbers of ATMs, number of credit and debit card and related volume of transactions only in 2011. The rapid speed of the spread of electronic banking can be judged by the fact that the total number of ATMs (online and offsite) across all banks increased from 75,645 to 160,055 between April 2011 and March 2014 amounting to a more than 100% increase. Comparable increases are found for credit and debit card transactions. In light of the significant degree of

¹ The Reserve Bank of India has continued a multi-pronged approach to achieve the goal of universal financial inclusion. Though that goal is still far away, marked progress has been made in recent years. For example, the number of banking outlets (branches and branchless mode) in villages went up from 67,694 in March 2010 to 586,307 in March 2016. Further, the number of basic savings bank deposit accounts have gone up from 73 million in March 2010 to 469 million in March 2016. (Reserve Bank of India Annual Report 2015-16.)

computerization in bank branches, the tasks performed by different categories of labor have undergone major changes. This has serious implications for labor cost efficiency of bank branches.

Since the reforms, several studies have focused on estimating the efficiency and productivity of Indian banks using both parametric and non-parametric methods (predominantly data envelopment analysis, (DEA)). However, from the perspective of a bank, achieving operational efficiency at the branch level is a prerequisite for the goals of technical and economic efficiency of the bank (Berger et al., 1997). In the U.S. and European banking context, therefore, a sizeable number of studies have examined the various dimensions of branch level efficiency (Paradi and Zhu, 2013). By contrast, to our knowledge, so far only two studies (Das et al., 2009 and Ray, 2016) have examined the efficiency of Indian banks at the branch level.

The analytical format for measuring efficiency of banks at the institutional level fits poorly when applied at the level of a branch of the bank. A bank primarily functions as a financial intermediary between borrowers and lenders of funds, transforming deposits and other borrowed funds into credit and investments. It is generally accepted in the literature that while, the *intermediation function* applies more appropriately to a bank, such functions are not totally at the discretion of individual branches. By contrast, the primary business of an individual branch is to provide deposit and credit services to the customers of the bank. At the branch level, labor and physical capital (like the premises and machines) serve as inputs. Services are indirectly measured by the number and type of transactions carried out for both deposits and loans, documents processed for taking deposits, extending loans and other direct services and so on. Hence, the *production approach* (introduced by Benston, 1965) is more appropriate for modeling the business of a bank branch wherein it is considered as a producer of services (Fethi and Pasiouras, 2010). Branches are responsible for managing their labor and capital efficiently in servicing the deposit and loan accounts of customers and offering other fee based services. Insights obtained from benchmarking analysis can help branches to continuously improve their operations and realign their businesses with the best practices available (Paradi and Zhu, 2013). In a typical bank branch in India, the extent and type of labor used by a branch depends not only on the volume of transactions provided but also on the *product mix* of a branch.

Our paper adds to the limited body of literature on operational efficiency of branches in the context of Indian banking. In this study we focus on a major public sector Indian bank with nationwide presence and examine the efficiency of 325 branches across four large metropolitan cities, Chennai, Delhi, Kolkata, and Mumbai. By including the branches of a single large bank, we are able to ensure that the same input prices apply to all branches, an important condition for the cost based formulation of the technology introduced in this paper. We utilize DEA to analyze data for the financial years 2008 and 2014. The year 2008 was a remarkable year for India with respect to banking development, growth, and

stability. During this year, both deposits and credit grew at an annual rate of 22 percent. Further, non-performing loans were at a historical low of around 1%.² Finally, this is the latest year of data prior to the global financial crisis. By 2014, however, the Indian economy had faced several challenges from the global financial crisis. The annual GDP growth was only 4.74% in 2013-14, marginally up from 4.47% in the previous year. According to the RBI, the growth of the banking sector had moderated further during the 2014 financial year and bank profitability had also declined due to stagnated credit growth as well as higher provisioning for delinquent loans.

This paper makes several contributions to the existing literature – both methodological and empirical. Relying on the neoclassical duality theory in production economics, we introduce a comprehensive analytical framework that addresses the question of efficiency measurement of bank branch operations, especially in the context of Indian banks.

- First, we propose a representation of the production technology in the form of a *cost set* in the output and expenditure space as an alternative to the standard free disposal convex hull of input-output vectors. We demonstrate that the disposability and convexity properties of the usual production possibility set carry over to the cost set as well.
- Second, we extend on the earlier work of Banker et al. (2007) and Ray and Mukherjee (2017) and demonstrate that when all units in the sample face identical input price vectors, the cost efficiency of any unit can be measured using only the data for output quantities and expenditures without any information on input quantities (or even input prices).
- Third, we show in this context that in the case of identical input prices across units, using input-specific expenditure data one can directly determine whether any unit is spending too much or too little on (i.e., over-utilizing or under-utilizing) any particular input. As we argue, our proposed measure of specific labor utilization rate is likely to be more useful for the decision maker than a scalar measure of allocative efficiency.
- As an empirical contribution, this is the first study in the Indian banking context to model the operations of branches using the *production approach* described above.
- Moreover, we impose integer constraints on the outputs measured by number of accounts.³
- Further, we control for the *product mix* in terms of the composition of loans, differentiating between large and small loan accounts.
- We also control for the average balance on deposit and credit accounts to ensure that in the optimal solution there is no reduction in the number of customers served by the branch.

² Source: Reserve Bank of India.

³ Das et al. (2009) considered indivisibility for the labor inputs only.

- Our choice of the two years for analysis allows us to compare the labor use efficiency of the same bank branches at an early and a more mature stage of computerization in India's public sector banks.
- Lastly, this study utilizes 650 observations for 325 branches across two years – 2008 and 2014, the latest year for which detailed branch level data were available. This is a larger sample size compared to most studies on bank branch efficiency. For instance, only a handful of the reported 80 studies in the survey paper by Paradi and Zhu (2013) use data sets containing more than 200 branches. Our use of a bigger data set should make our empirical findings more reliable.

Overall, we believe that the main contributions of this paper lie in both extending the methodology and providing an important empirical application in the context of branch banking.⁴ Our empirical findings indicate that there is significant inefficiency in labor use in the branches and cost could be curtailed substantially by addressing overstaffing. Across the three types of labor, reducing the expenditure of clerks would have the highest impact for cost saving. We do find, however, that the extent of overspending on clerks has reduced in 2014, which apparently is a direct consequence of computerization of routine jobs. Efficiency varies across regions. In general Chennai branches are more efficient than branches from other regions whereas Kolkata branches are the least efficient.

The rest of the paper is organized as follows. In Section 2, we provide a brief review of the literature on efficiency in branch banking. In Section 3, we explain the non-parametric methodology and the proposed extensions. Section 4 describes the data construction and modeling issues and presents the cost-based dual models used in our empirical application. Section 5 provides a discussion of the main results and Section 6 concludes.

2. Review of Literature

The operational efficiency of banks has been a matter of immense and growing interest in all countries. However, three major survey papers on bank performance by Berger and Humphrey (1997), Berger (2007), and Fethi and Pasiouras (2010) all reveal that a majority of papers focus on performance of a bank at the institutional level. A much smaller number of studies address performance at the branch level using either parametric or non-parametric methods⁵. A recent survey paper by Paradi and Zhu (2013) also finds that of the 275 studies between 1985-2011 that utilized DEA for examining bank sector performance, only 80 were focused on efficiency at the branch level. The first application of bank branch

⁴ Other notable studies that have made methodological contributions in the context of bank branch performance include Athanassopoulos (1997), Camanho and Dyson (1999), Drekkker and Post (2001), Wu et al. (2006), and Paradi et al. (2010).

⁵ A recent study by Quaranta et al. (2018) proposes a multidimensional approach to measuring branch bank efficiency.

efficiency analysis was by Sherman and Gold (1985). Since then, the number of branch studies has increased over time, with 53 of the 80 studies being conducted in 2001 and later. Further, of the 80 studies, 65% were focused on the U.S. and major European countries. One reason for the limited number of branch level studies in other countries is the lack of publicly available reliable data at this level of operation.

There is now a sizeable literature examining the efficiency and productivity of banks in the Indian banking sector, especially since the 1990s financial reforms. While Saha and Ravisankar (2000) focuses only on public sector banks, a major aspect of analysis of most studies has been the comparison of performance of banks across different ownership types (Sathye, 2003; Bhaumik and Dimova, 2004; Ram Mohan and Ray, 2004; Sensarma, 2006; Das and Ghosh, 2006; Zhao et al., 2010; and Ray and Das, 2010, among others). Recent studies along these lines include Casu et al. (2013), Fujii et al. (2014), Tzeremes (2015) and Badunenko and Kumbhakar (2017).

The literature on efficiency and productivity of Indian bank branches, however, is very limited. To our knowledge, only two published studies have addressed this aspect of bank performance (Das et al., (2009) and Ray (2016)). Das et al., (2009) use DEA to measure labor-use efficiency of 222 individual branches of a large public sector bank in 2003 across four metropolitan cities. They model the operation of a branch using the *intermediation approach*. Using a traditional DEA model of variable cost minimization model, they find considerable variation in efficiency of bank branches across the metropolitan regions which indicate that the effects of social and regional differences are important in spite of the same organizational set up.

Ray (2016) takes a very different perspective and evaluates the ‘overall’ cost efficiency of a network of 193 branches of a single large public sector bank within the city of Kolkata using data for the year 2012. The study attempts to determine the optimal number of branches within a postal district that could provide the observed amounts of banking services to customers in that area at the minimum operating cost. The DEA results show that overall there is evidence of ‘over-branching’. However, there are also instances where increasing the number of branches within a market area would be optimal.

To summarize, the current study revisits the estimation of labor cost efficiency in Indian bank branches with a detailed data set pertaining to the year 2008, the final year of what is often described as India’s short lived era of ‘golden growth’ as well as the year 2014 which is characterized as the first year of the neo liberal regime under the current Modi government. This permits us to offer new and valuable insights into the performance of the Indian banking industry at a disaggregated branch level. To the extent that the branch banking scenario in India, characterized by heterogeneous product mix, can be generalized to other large emerging economies, our modeling structure would be appropriate for other studies as well.

3. The Nonparametric Methodology

For efficiency measurement using econometric techniques, one specifies some explicit form of the production, cost, or profit function to represent the benchmark technology. In the nonparametric alternative, however, one makes a number of fairly general assumptions about the technology and leaves the functional form unspecified. Typically, it is assumed that the production possibility set is convex and both inputs and outputs are freely disposable.

Consider an industry producing m outputs from n inputs. An input-output bundle $(x, y): x \in R_+^n, y \in R_+^m$ is considered feasible when the output bundle y can be produced from the input bundle x . The technology faced by the firms in the industry can be described by the production possibility set

$$T = \{(x, y): y \text{ can be produced from } x\} \quad (1)$$

It is conventional to assume that

(A1) every observed input-output bundle is feasible; (A2) inputs are freely disposable; (A3) outputs are freely disposable; (A4) the production possibility set is convex.

Suppose that the data set $D = \{(x^j, y^j); j = 1, 2, \dots, N\}$ consists of observed input-output vectors (x, y) from N firms in the industry,

Then

(1) $(x^j, y^j) \in D \Rightarrow (x^j, y^j) \in T$ by (A1);

(2) $(x^0, y^0) \in T \wedge x^1 \geq x^0 \Rightarrow (x^1, y^0) \in T$ by (A2).

(3) $(x^0, y^0) \in T \wedge y^1 \leq y^0 \Rightarrow (x^0, y^1) \in T$ by (A3).

(4) By the convexity assumption (A4)

$$(x^1, y^1), (x^2, y^2) \in T \wedge 0 \leq \lambda \leq 1 \Rightarrow (\lambda x^1 + (1 - \lambda)x^2; \lambda y^1 + (1 - \lambda)y^2) \in T$$

Under the assumptions (A1) - (A4) an empirical estimate of the production possibility set is the free disposal convex hull of the observed data points

$$S = \{(x, y): x \geq \sum_{j=1}^N \lambda_j x^j; y \leq \sum_{j=1}^N \lambda_j y^j; \sum_{j=1}^N \lambda_j = 1; \lambda_j \geq 0; (j = 1, 2, \dots, N)\} \quad (2)$$

3.1 Alternative Measures of Efficiency

The two most popular measures of efficiency in the DEA literature are the radial input and output oriented measures of technical efficiency due to Farrell (1957), Charnes, Cooper, and Rhodes (CCR)(1978), and Banker, Charnes, and Cooper (BCC)(1984). In an empirical analysis, for any unit using input x^0 to produce output y^0 the radial input-oriented efficiency is

$$\tau_x(x^0, y^0) = \min \theta: (\theta x^0, y^0) \in S \quad (3)$$

and the radial output-oriented efficiency is

$$\tau_y(x^0, y^0) = \frac{1}{\varphi^*} \text{ where } \varphi^* = \max \varphi: (x^0, \varphi y^0) \in S \quad (4)$$

When output is preassigned as a given task, efficiency lies in delivering the target output using the minimum quantity of inputs. In this respect, however, the radial input-oriented measure in (3) is rather inadequate because it does not capture the potential for reducing individual inputs to their fullest extent. A more complete measure of input-oriented technical efficiency is the non-radial Russell measure:

$$\rho(x^0, y^0) = \min \frac{1}{n} \text{trace}(\Theta): (\Theta x^0, y^0) \in S; 0 \leq \Theta \leq I \quad (5)$$

where Θ is a diagonal matrix, O is the null and I is the identity matrix.

Although better than the radial measure, the Russell measure is still somewhat unsatisfactory because it treats reduction in all inputs equally without any reference to their market prices. When reliable price information is available, economizing on the use of inputs amounts to minimizing the total cost of the input bundle used even if it might require increasing some individual inputs while reducing others. At this point, it is useful to consider the empirically constructed input requirement set⁶ of the observed output bundle y^0

$$\hat{V}(y^0) = \{x: (x, y^0) \in S\} \quad (6)$$

The minimum cost of producing y^0 at input prices w^0 is

$$C(w^0, y^0) = \min w^{0'} x: x \in \hat{V}(y^0) \quad (7)$$

Following Farrell (1957), the cost efficiency of the unit can be measured by the ratio of the minimum cost to its actual cost

$$\gamma(w^0, y^0; C^0) = \frac{C(w^0, y^0)}{C^0} \text{ where } C^0 = w^{0'} x^0. \quad (8)$$

We next introduce a cost-aggregated representation of the input-requirement set which we define as a cost set and highlight some of its properties which follow from the convexity and disposability assumptions A1 through A4 above.

3.2 A Cost Set Representation of the Technology

For a given vector of input prices w^0 , one can evaluate the cost $e = w^{0'} x$ of every input bundle $x \in V(y^0)$. Because $C(w^0, y) \leq w^{0'} x \forall x \in V(y)$, a dual representation of the overall technology through its *cost set* is

$$E(e, y|w^0) = \{(e, y): e \geq C(w^0, y)\}. \quad (9)$$

It can be easily shown that

$$(a) (e_0, y) \in E(e, y|w^0) \wedge e_1 > e_0 \Rightarrow (e_1, y) \in E(e, y|w^0).$$

Clearly, $(e_0, y) \in E(e, y|w^0)$ implies $e_0 \geq C(w^0, y)$. Hence, $e_1 > e_0 \geq C(w^0, y)$. Thus, $e_1 \in E(e, y|w^0)$.

⁶ An input requirement set consists of all the input bundles that can produce the specified output.

(b) $(e, y^0) \in E(e, y|w^0) \wedge y^1 \leq y^0 \Rightarrow (e, y^1) \in E(e, y|w^0)$. This follows from free disposability of outputs, which implies that if $x \in V(y^0)$, then $x \in V(y^1)$ for any $y^1 \leq y^0$.

(c) $E(e, y|w^0)$ is convex. Consider the output bundles y^0 and y^1 . Suppose x^{0*} minimizes the cost of producing y^0 and x^{1*} minimizes the cost of producing y^1 at input prices w^0 . Now consider the output bundle $\bar{y} = \lambda y^0 + (1 - \lambda)y^1$ ($0 < \lambda < 1$). By convexity of the technology, the input bundle $\bar{x} = \lambda x^{0*} + (1 - \lambda)x^{1*} \in V(\bar{y})$. However $\bar{x}$ may not be cost minimizing for $\bar{y}$. Hence, $\bar{e} = w^{0'}\bar{x} \geq C(w^0, \bar{y})$ implying that $\bar{e} \in E(e, \bar{y}|w^0)$. Now, $e_0 \in E(e, y^0|w^0) \Rightarrow e_0 \geq C(w^0, y^0) = w^{0'}x^{0*}$. Similarly, $e_1 \in E(e, y^1|w^0) \Rightarrow e_1 \geq C(w^0, y^1) = w^{0'}x^{1*}$.

Therefore, $\bar{e} = \lambda e_0 + (1 - \lambda)e_1 \geq w^{0'}(\lambda x^{0*} + (1 - \lambda)x^{1*}) \geq C(w^0, \bar{y})$
 $\Rightarrow \bar{e} = \lambda e_0 + (1 - \lambda)e_1 \in E(e, \bar{y}|w^0)$.

The dual cost function $e = C(w^0, y)$ is the lower boundary or the frontier of the cost set for a given vector of input prices.

3.3 Cost Minimization and the Dual Approach

In stochastic frontier analysis (SFA) one needs the data for only the output quantities and input prices in order to estimate the dual cost frontier. In particular, one does not need information on the input quantities. This turns out to be especially helpful in situations where input quantity data are unavailable. By contrast, in order to solve the cost minimization problem in DEA models one requires the construction of the free disposal convex hull of input-output vectors as the estimated production possibility set. This obviously requires information on the input quantities. In this regard, DEA is at a considerable disadvantage relative to SFA. Here we show that the input quantity data requirement can be waived even for DEA if all firms in the sample face the same input price vector (w). Consider the cost minimization problem

$$C(w, y) = \min \sum_{i=1}^n w_i x_i \quad \text{s.t.} \quad (10)$$

$$\sum_{j=1}^N \lambda_j x_{ij} \leq x_i, (i = 1, 2, \dots, n);$$

$$\sum_{j=1}^N \lambda_j y_{rj} \geq y_r^0, (r = 1, 2, \dots, m);$$

$$\sum_{j=1}^N \lambda_j = 1; \lambda_j \geq 0, (j = 1, 2, \dots, N).$$

Here x_{ij} is the observed quantity of input i used and y_{rj} is the observed output r produced by unit j and y_r^0 is the actual quantity of output r produced by the unit under evaluation. Also, note that w_i , the price of input i , is applicable to all units. Now, multiplying both sides of each input constraint by the relevant input price and adding over all inputs, we get

$$\begin{aligned}
C(w, y) &= \min \sum_{i=1}^n w_i x_i \\
\text{s.t.} & \\
\sum_{j=1}^N \lambda_j \left(\sum_{i=1}^n w_i x_{ij} \right) &\leq \sum_{i=1}^n w_i x_i; \\
\sum_{j=1}^N \lambda_j y_{rj} &\geq y_r^0, (r = 1, 2, \dots, m); \\
\sum_{j=1}^N \lambda_j &= 1; \lambda_j \geq 0, (j = 1, 2, \dots, N).
\end{aligned} \tag{11}$$

However, $\sum_{i=1}^n w_i x_{ij} = C_j^0$, the actual cost of unit j and $\sum_{i=1}^n w_i x_i \equiv C$ is the objective function. Hence, the cost minimization problem, in this case, becomes

$$\begin{aligned}
C(w, y) &= \min C \\
\text{s. t.} & \\
\sum_{j=1}^N \lambda_j C_j^0 &\leq C; \\
\sum_{j=1}^N \lambda_j y_{rj} &\geq y_r^0, (r = 1, 2, \dots, m); \\
\sum_{j=1}^N \lambda_j &= 1; \lambda_j \geq 0, (j = 1, 2, \dots, N).
\end{aligned} \tag{12}$$

One can now use for y^0 the output vector of any individual unit k in the data set to solve the above problem and measure its cost efficiency as $\frac{C(w, y^k)}{C_k^0}$.

It needs to be emphasized that, as noted in Banker et al. (2007) and Ray and Mukherjee (2017), the DEA LP in (12) using outputs and expenditure data yields a valid measure of cost efficiency only if all firms face identical input price vectors.⁷

Two points need to be emphasized with respect to the relation between the disaggregated and aggregated models. First, while the transition from the disaggregated model in (10) to the aggregated model in (12) is quite transparent, there is no obvious way to arrive at the disaggregated model starting from the aggregated model. We show below, however, that the optimal solution to the aggregated model will invariably be the same as the optimal solution to the disaggregated model.

An important point to note is that in all of the models (10) - (12), each input constraints must hold as strict equalities at the optimal solution. In other words, there can be no input slacks in a cost minimization problem. Suppose that in the disaggregated model (10), the optimal solution is

$(\lambda_j^*; x_i^*; j = 1, 2, \dots, N; i = 1, 2, \dots, n)$ and $C^* = \sum_{i=1}^n w_i x_i^*$. Then for each input i , $\sum_{j=1}^N \lambda_j^* x_{ij} = x_i^*$, and $\sum_{j=1}^N \lambda_j^* (\sum_{i=1}^n w_i x_{ij}) = \sum_{i=1}^n w_i x_i^*$. That is $\sum_{j=1}^N \lambda_j^* C_j^0 = C^*$. Because the optimal solution from (10)

is a feasible solution for (12), the minimum from the aggregated model in (12) cannot be any higher than what we get in (10). Next, consider the problem in (12). This time suppose that the optimal solution is

$(\lambda_j^{**}; j = 1, 2, \dots, N)$ and $\sum_{j=1}^N \lambda_j^{**} C_j^0 = C^{**}$. Recall, however, that

$$\sum_{j=1}^N \lambda_j^{**} C_j^0 = \sum_{j=1}^N \lambda_j^{**} (\sum_{i=1}^n w_i x_{ij}) = \sum_{i=1}^n w_i (\sum_{j=1}^N \lambda_j^{**} x_{ij}) = C^{**}.$$

Now define $\sum_{j=1}^N \lambda_j^{**} x_{ij} = x_i^{**}, (i = 1, 2, \dots, n)$. Then $(\lambda_j^{**}; x_i^{**}; j = 1, 2, \dots, N; i = 1, 2, \dots, n)$ is a feasible solution for (10) and the optimal solution cannot be any higher than C^{**} . This completes the proof that both the disaggregated model (10) and the aggregated model (12) yield the same optimal value of the objective function. An implication of this is that so long as all units face the same input price vector, data on outputs and total expenditures are sufficient for measuring cost efficiency. One does not need to know the actual input prices. Further, availability of individual input quantity data would not lead to a different value of the objective function than what we get with total expenditure data only.

This aggregated model in (12) to measure cost efficiency does not provide any information about allocative efficiency. One also needs a measure of technical efficiency to derive a measure of allocative efficiency from overall cost efficiency. However, if in addition to the information on total expenditure, one also knows its input-wise break down, it is possible to determine whether a firm is spending too much

⁷ Some studies use total expenditure as a measure of aggregate input, which is clearly inappropriate if input prices are not identical across units. Moreover, even when input prices are identical across the sample, this is model measures *cost efficiency* and not *technical efficiency* (unless all firms are assumed to be allocatively efficient). In the present application, however, all branches of the bank under consideration do pay the same input prices.

or too little on individual inputs. Let $c_{ij} = w_i x_{ij}$ be the observed expenditure on the input i by the j th firm. For example, it may be the amount spent on labor by the firm. While this is definitely an additional information, one may know it without knowing either the number of workers or the wage rate. The model (11) can, then, be expressed as

$$\begin{aligned} C(w, y) = \min \sum_{i=1}^n c_i \\ \text{s. t.} \end{aligned} \quad (13)$$

$$\begin{aligned} \sum_{j=1}^N \lambda_j c_{ij} &\leq c_i; \quad (i = 1, 2, \dots, n); \\ \sum_{j=1}^N \lambda_j y_{rj} &\geq y_r^0, \quad (r = 1, 2, \dots, m); \\ \sum_{j=1}^N \lambda_j &= 1; \quad \lambda_j \geq 0, \quad (j = 1, 2, \dots, N). \end{aligned}$$

In the model above, at the optimal solution c_i^* would denote the cost minimizing expenditure on the i^{th} input⁸. If the actual expenditure is higher (lower) than this optimal expenditure, the firm is using too much (too little) of the input i . In a sense, such information about over or under use of individual inputs is more useful than a summary measure of allocative efficiency.

Suppose c_i^0 is the actual expenditure on input i by the unit under evaluation (identified by the superscript 0). One can define the i^{th} input utilization rate as⁹

$$\eta_i^0 = \frac{c_i^0}{c_i^*}. \quad (14)$$

When c_i^0 exceeds (falls short of) c_i^* , the unit is over (under) utilizing the i^{th} input. For $\eta_i^0=1$, the actual expenditure on the i^{th} input is optimal.

3.4 Group Specific and Industry Cost Frontiers

In the production possibility set (2) represented by the free disposal convex hull of all of the observed input-output data points in the sample, it was implicitly assumed that every decision making unit has access to all points in the technology set. In many cases there are geographical and/or

⁸ Note that we are using upper case C to denote total cost of labor of a branch (as in (12)) and lower case c to denote the cost of individual labor category (as in (13)).

⁹ This is comparable to the definition of variable input utilization rate by Färe et al.(1994) in their measurement of capacity utilization based on the production frontier.

institutional factors specific to units from different groups that introduce a degree of technological heterogeneity. In that case, one can think of group-specific technology sets subsumed under the overall or industry-wide production possibility set shown in (2). Intuitively, this implies that while input-output bundles are interchangeable across and/or may be combined with units within a group, they may not be so between groups.

The concept of a *meta frontier* encompassing many *group frontiers* was introduced by Hayami and Ruttan (1971). Subsequently Battese and Rao (2002) extracted three separate factors representing *technology gap ratio (TGR)*, *technical efficiency ratio (TER)*, and a *random error ratio (RER)* from the ratio of the estimated *group* and *meta* frontiers at the observed data points. In fact, the concept of *technology gap ratio (TGR)* has been used in subsequent studies including Battese et al. (2004), Bhandari and Ray (2011) among many others. Note that all previous *meta frontier* studies have used the production frontier. By contrast, this paper uses the dual cost representation of the technology set described above to construct an *industry cost frontier* and several *group specific cost frontiers*. We use the group frontier to measure how well the individual units within a group performed relative to a frontier constructed from data pertaining to other units from that group only. Given that our groups represent different metro areas, our *area efficiency*, as defined below, is a summary measure of the proximity of the group frontier to the industry (or meta) frontier and is a weighted average of the technology gap ratios for the individual units in a group.

Let the set J consist of all units in the data. Partition the entire set into G groups as $J = \{J_1, J_2, \dots, J_G\}$. Any unit $j \in J_g$ if and only if it belongs to group g . We may now create group-specific sub technologies for each group $g = (1, 2, \dots, G)$.

$$S_g = \left\{ (x, y) : x \geq \sum_{j \in J_g} \lambda_j x^j, y \leq \sum_{j \in J_g} \lambda_j y^j, \sum_{j \in J_g} \lambda_j = 1; \lambda_j \geq 0; j \in J_g \right\} \quad (15)$$

The corresponding input requirement sets are

$$V_g(y^0) = x : (x, y^0) \in S_g \quad (16)$$

resulting in group-specific minimum cost

$$C_g(w^0, y^0) = \min w^0 x : x \in V_g(y^0) \quad (17)$$

and cost efficiency

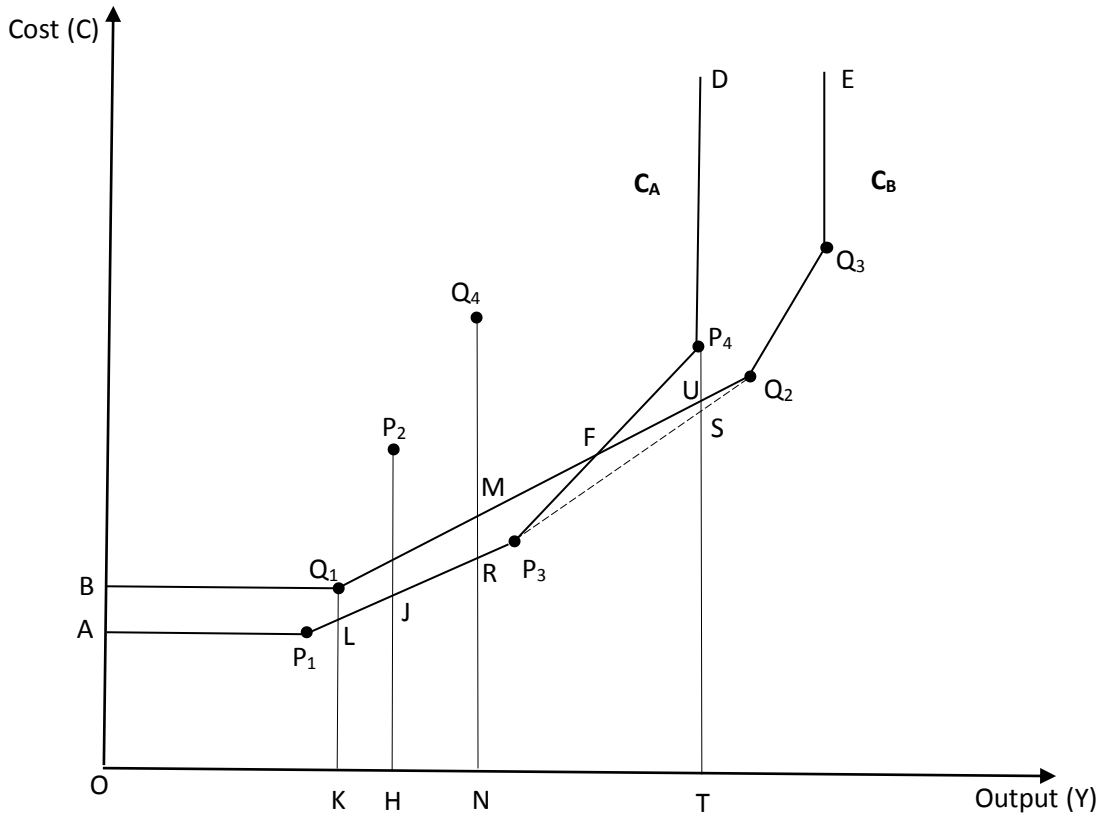
$$\gamma_g(w^0, y^0; C^0) = \frac{C_g(w^0, y^0)}{C^0} \text{ where } C^0 = w^0 x^0. \quad (18)$$

For each unit $j \in J_g$, one can now compute two different measures of cost efficiency. The first, $C_g(w^0, y^j)$ is relative to the group frontier. The other, $C(w^0, y^j)$ is relative to the industry (or grand) frontier. The first measures how close the unit is to the group frontier. The second, on the other hand, measures, how

close the unit is to the industry frontier. The ratio of the two is a measure of the proximity of the group frontier to the industry frontier evaluated at a specific output level.¹⁰

This is illustrated geometrically in Figure 1. Suppose that the units can be categorized into two groups. Assume, for simplicity, that all firms face the same input prices. The point P_1 through P_4 show the output quantities and the actual costs of the 4 firms from group 1. The cost frontier for this group is shown by the piecewise linear lower envelop $AP_1P_3P_4D$. Similarly, points Q_1 through Q_4 show the outputs and costs of the firms from group 2 and the corresponding cost frontier is $BQ_1Q_2Q_3E$. The overall (industry) frontier is the outer envelope of the two group frontiers and is shown by the broken and piecewise connected line $AP_1P_3Q_2Q_3E$ corresponding to the technology set defined in (2). Note that a segment of this frontier – identified by the broken line P_3Q_2 lies below both of the group frontiers. A point in this interval corresponds to a convex combination of the output bundles of firm 3 in group 1 and firm 2 in group 2 and the cost of the corresponding convex combination of their input bundles. This, it may be noted, is feasible in case of the overall frontier but not for the individual group frontiers.

Figure 1: Group-Specific and Industry Cost Frontiers



¹⁰ Battesse and Rao (2002) characterize this as the ‘technology gap’ in the context of a production frontier.

Consider, now, firm 1 from group 2 producing output OK at the cost Q_1K . The point Q_1 is on the cost frontier for group 2 and it is efficient within the group. However, the minimum cost for the same output as per the overall industry cost frontier is LK and this unit is not efficient relative to the industry frontier. The ratio $\frac{LK}{Q_1K}$ is a measure of the cost efficiency of group 2 relative to the entire industry evaluated at the output level OK . For firm 4 in group 2 producing the output ON , the actual cost is Q_4N while the relevant points on the group and industry frontier are M and R , respectively. Thus, the group specific cost efficiency of firm 4 in group 2 is $\frac{MN}{Q_4N}$ while the efficiency of group 2 at the output level ON is $\frac{RN}{MN}$. In the case of firm 4 from group 1, point P_4 is located on the group frontier showing a 100% group specific efficiency for this unit. The cost of the same output level OT based on the group 2 frontier would be UT which is lower than P_4T . But the corresponding benchmark point on the industry frontier is S which is even lower being located on the P_3Q_2 segment connecting two points from the different frontiers.

A measure of the average cost efficiency of the units within the group relative to its group frontier is

$$\bar{\gamma}_g = \frac{\sum_{j \in J_g} C_g(w^0, y^j)}{\sum_{j \in J_g} w^{0'} x^j} = \sum_{j \in J_g} \left(\frac{w^{0'} x^j}{\sum_{j \in J_g} w^{0'} x^j} \right) \left(\frac{C_g(w^0, y^j)}{w^{0'} x^j} \right) = \sum_{j \in J_g} \alpha_j \gamma_g(w^0, y^j), \quad (19)$$

$$\text{where } \alpha_j = \frac{w^{0'} x^j}{\sum_{j \in J_g} w^{0'} x^j} \quad (20)$$

is the share of unit j in the total cost of all units in group g . A proportional measure of the cost saving that could be achieved if all units could be made efficient relative to the group cost frontier is $(1 - \bar{\gamma}_g)$.

Similarly, the average cost efficiency of the units within the group corresponding to the industry frontier is

$$\bar{\bar{\gamma}}_g = \frac{\sum_{j \in J_g} C(w^0, y^j)}{\sum_{j \in J_g} w^{0'} x^j}, \quad (20a)$$

which is also a cost share weighted average of the efficiencies of units in the group.

Hence, a measure of the overall efficiency of group g relative to the entire industry is

$$\mu_g = \frac{\sum_{j \in J_g} C(w^0, y^j)}{\sum_{j \in J_g} C_g(w^0, y^j)} = \frac{\bar{\bar{\gamma}}_g}{\bar{\gamma}_g} \quad (21)$$

$$\bar{\bar{\gamma}}_g = (\bar{\gamma}_g) \cdot (\mu_g) \quad (21a)$$

A low value of μ_g indicates that the group cost frontier of group g is located much above the industry cost frontier and $(1 - \mu_g)$ is a measure of the proportional decrease in the total cost if all firms from group g could move from the group frontier to the industry frontier. While $\gamma_{gj} = (w^0, y^j)$ measures the performance of the decision maker at unit j relative to its own group, μ_g is an overall measure of the performance of the group relative to the entire industry. When the groups correspond to geographical

areas, the group efficiency can be interpreted as ‘area efficiency’.¹¹ While previous studies such as Das et al (2009) do measure area efficiency, their use of unweighted averages do not accurately reflect the relevant potential reduction in cost.

3.5 Indivisibility of Inputs and/or Outputs

We have so far assumed that all of the inputs and outputs are non-negative, real-valued, continuous variables. Suppose, however, that some of these inputs and outputs can only be positive integers. This would immediately invalidate both the convexity and disposability assumptions in (A2)-(A4). Even when x_1 and x_2 are integer valued, every convex combination $\bar{x} = \lambda x_1 + (1 - \lambda)x_2$; ($0 \leq \lambda \leq 1$) need not also be integer valued. In the same vein, even when (x_1, y_1) is feasible, $\delta > 0$ does not imply $(x_1 + \delta, y_1)$ is also feasible, unless δ is also an integer. Similar reasoning applies about the output when y also has to be an integer.

Kuosmanen and Matin (2009) and Matin and Kuosmanen (2009) introduced the concepts of *Integrality*, *Natural Disposability*, and *Natural Convexity* in the context of DEA in order to accommodate integer outputs and inputs.¹² Consider the most general case where some inputs and outputs are integer valued while the others can take any non-negative real value.

Accordingly, partition the input vector as $x = (x^I \mid x^{II})$ where $x^I \in R_+^{n_1}$ is restricted only to be real valued whereas $x^{II} \in Z_+^{n_2}$ must also be integer valued.¹³ Similarly, partition $y = (y^I \mid y^{II})$ where $y^I \in R_+^{m_1}$ and $y^{II} \in Z_+^{m_2}$. Of course, when $n_2 = 0$ all inputs can take any non-negative real value and for $n_2 = m_2 = 0$ we get back to the usual DEA models.

As discussed in section 4, in our empirical application, several of our outputs are indivisible and our model (22) explicitly incorporates integer constraints for these outputs. However, since we use different categories of expenditures as our inputs, indivisibility does not apply to our inputs.

¹¹ In our empirical application, there are four groups ($g = 1, \dots, 4$), pertaining to each metro Chennai, Delhi, Kolkata, and Mumbai. Accordingly, the group efficiency (μ_g) of a metro with respect to the industry frontier is the ‘area efficiency’ of that specific metro.

¹² Integrality implies the relevant input or output can only be integer valued. Natural Disposability means that if any input-output bundle is feasible, reducing any output or increasing any input by an integer amount does not affect the feasibility of the new bundle. Similarly, Natural Divisibility means that one can scale down any feasible input-output bundle to a smaller bundle where the reduced input-output quantities are integers. Finally, a set is Natural Convex when convex combinations of its elements belong to the set only if they also are integers.

¹³ In manufacturing, for example, production labor measured in hours can take fractional values but the number of managers must be an integer.

4. The Empirical Application

4.1 Data Construction and Modeling Issues

Our study focuses on the labor cost efficiency of branches of a large public sector bank in India with nationwide presence. Data relate to the financial years 2008 (April 2007 to March 2008) and 2014 (April 2013 to March 2014). Branches of Indian banks are engaged in one or more of the following functions: (i) banking business (i.e., accepting deposit and /or offering credit to their customers), (ii) administration, and (iii) foreign exchange business. In this analysis we include only retail branches that focus on consumer banking. As in other countries, operations of a typical bank branch in India involve collection of deposits, making loans and providing other financial services. However, due to location (dis)advantage, other socio-economic characteristics of customers as well as strategic decisions, the primary function of a large number of bank branches in India is deposit mobilization. Similarly, there are branches whose major focus is offering credit - sometimes only to a particular segment of industry. As a result, it is quite common to observe very high or low credit-deposit ratio (C/D ratio) of a branch in the same city. In choosing our sample we excluded branches with zero or missing values of inputs and outputs. Further, following the earlier literature (Das et al., 2009, Ray, 2016), we excluded very specialized branches with extremely high or low C/D ratios. We included in our sample only those branches with C/D ratio between 15% at the lower end and 85% at the upper end in 2008 and evaluated these same branches for 2008 and 2014. Our sample comprises 325 branches, with 72 branches in Chennai, 67 branches in Delhi, 123 branches in Kolkata, and 63 branches in Mumbai.

In conceptualizing the operations of a branch we follow the *production approach*¹⁴ and assume that a branch uses three types of employees - (i) officers, (ii) clerks, and (iii) subordinates, along with (iv) capital (measured by overhead expenses) to produce four outputs - (i) number of large loan accounts (with more than ₹200,000 balance), (ii) number of small loan accounts (with less than ₹200,000),¹⁵ (iii) number of deposit accounts, and (iv) non-interest income (as a proxy for other fee based services offered by a branch).¹⁶ Given that the fee structure of this bank is the same across the branches, the monetary value of this service is an acceptable measure. Ideally, the number of loan and deposit *transactions* (rather than

¹⁴ As explained previously, ‘production approach’ and ‘intermediation approach’ are two alternative approaches for defining the inputs and outputs of a bank or branch. Subsequent to this, either a primal production frontier analysis or a dual cost frontier analysis can be applied for either of these two approaches with the appropriate definition of inputs and outputs.

¹⁵ The Bank in our study used ₹200,000 (in current Rupees) as the cutoff point to differentiate between its large and small loan accounts during our sample years. It is worthwhile to note that this cutoff figure used by the bank was not adjusted for inflation.

¹⁶ ₹ is the currency symbol for the Indian Rupee (INR). The conversion rate was 1 USD = 48.2 INR as of December 31, 2008 and 63.2 INR as of December 31, 2014. Thus, the value of a small loan was less than approximately \$4,150 in 2008 and \$3,165 in 2014.

accounts) should be used as the outputs of a branch since resource use is determined by the number of transactions serviced (Sherman and Gold, 1985). However, such detailed data is not available and the number of accounts was the best possible way to measure the outputs. By explicitly including the two categories of loans, we account for the fact that origination and monitoring of larger loans are more important from risk and stability point of view, and requires a branch to devote more resources in terms of screening and credit check of the borrower and may also involve more frequent transactions. The wages of the different categories of labor are determined at the bank level and are the same across all branches. Accordingly, using model (13) we measure cost efficiency using the individual components of labor costs rather than units of labor as our inputs.

Table 1 presents the summary data for our sample of 325 branches. We see that in 2008, 69.82% of the operational cost (non-interest expenses) of an average branch was incurred on labor and the remaining 30.18% on overhead expenses. The ratio remained very similar in 2014 with the average branch spending 69.86% of its operational cost on labor and 30.14% on overhead expenses. It is interesting to examine the relative costs of the three types of labor. In 2008, the average branch spent 33.38% of the total labor cost on officers, 47.90% on clerks, and 18.72% on subordinates. In 2014, the average branch spent 47.82% of the total labor cost on officers, 39.42% on clerks, and 12.75% on subordinates. The increase in the relative share of cost on officers was both due to increase in the relative number of officers at a branch as well as the wages of officers. This is consistent with the findings from the impact of skill biased technical change that has resulted in the increased income share for people in decision making roles at the expense of those performing routine and repetitive jobs. For the average branch, the composition of the loan portfolio also changed across the two years with large loans accounting for 36.30% of the loan portfolio in 2008 but a 65.52% of the loan portfolio in 2014. The decline in the number of small loan accounts side by side an increase in the large loan accounts is only to be expected because the definition of large loan accounts has not been adjusted upwards to be consistent with inflation. The nominal value of average transactions have gone up due to inflation as well as increased purchasing power. The branches differed widely in terms of the number of loan and credit accounts (as seen from the minimum and maximum values), with the average branch managing roughly 13,161 deposit and 724 loan accounts in 2008 and 22,856 deposit and 839 loan accounts in 2014.

Table 1: Summary Statistics for the Data

Variable	2008				
	Mean	Median	SD	Min	Max
N=325					
Expenditure on Officers (₹'000)	3,178.21	2,716	2,031	679	13,582
Expenditure on Clerks(₹'000)	4,559.89	3,468	3,371	347	26,705
Expenditure on Subordinates (₹'000)	1,782.26	1,343	1,236	269	9,940
Overhead Expenses (₹'000)	4,115.37	3,324	2,773	422	17,120
No. of Large Loan Accounts	262.92	208	217	22	1,419
No. of Small Loan Accounts	461.43	345	432	4	3,545
No. of Deposit Accounts	13,160.89	11,391	8,034	330	54,724
Non-interest Income (₹'000)	5,032.44	2,795	7,708	99	80,714
Large Loan Account Balance (₹'000)	878.66	614	1,306	263	18,483
Small Loan Account Balance(₹'000)	58.47	57	12	19	136
Deposit Account Balance (₹'000)	67.84	44	115	15	1,827
Variable	2014				
	Mean	Median	SD	Min	Max
N=325					
Expenditure on Officers(₹'000)	7,143.94	6,645	4,381	1,329	27,909
Expenditure on Clerks (₹'000)	5,888.73	4,859	3,597	540	26,454
Expenditure on Subordinates (₹'000)	1,905.12	1,361	1,446	454	8,165
Overhead Expenses (₹'000)	6,445.25	5,035	5,226	1,028	49,566
No. of Large Loan Accounts	549.56	417	419	43	2,206
No. of Small Loan Accounts	289.18	202	283	34	2,318
No. of Deposit Accounts	22,855.88	19,846	12,465	3,169	71,997
Non-interest Income (₹'000)	9,338.50	5,469	14,770	525	172,083
Large Loan Account Balance (₹'000)	988.97	738	1,191	309	16,783
Small Loan Account Balance(₹'000)	63.64	62	23	23	361
Deposit Account Balance (₹'000)	112.97	88	98	2	853

Note: The nominal data for 2008 and 2014 are in current Rupees. Between the years 2008 and 2014, the GDP deflator (base year 2004-05 = 100) increased from 117.6 to 173.5.

4.2 The Proposed Labor Cost Minimization DEA Model

Given that the main objective of eliminating labor use inefficiency is to reduce the personnel costs, a proper optimization problem is one that minimizes the total wage payment. Accordingly, the DEA model (13) can be written in the context of our empirical application as:

$$\begin{aligned}
C^* &= \text{Min. } c_1 + c_2 + c_3, \text{ subject to} \\
\sum_j \lambda_j c_{1j} &\leq c_1 \text{ (Expenditure on officers);} \\
\sum_j \lambda_j c_{2j} &\leq c_2 \text{ (Expenditure on clerks);} \\
\sum_j \lambda_j c_{3j} &\leq c_3 \text{ (Expenditure on subordinates);} \\
\sum_j K_j &\leq K^0 \text{ (overhead expenses);} \\
\sum_j \lambda_j y_{1j} &= y_1; y_1 \geq y_1^0 \text{ (No. of large loan accounts);} \\
\sum_j \lambda_j y_{2j} &= y_2; y_2 \geq y_2^0 \text{ (No. of small loan accounts);} \\
\sum_j \lambda_j y_{3j} &= y_3; y_3 \geq y_3^0 \text{ (No. of deposit accounts);} \\
\sum_j \lambda_j y_{4j} &\geq y_4^0 \text{ (non – interest income);} \\
\sum_j \lambda_j z_{1j} &\geq z_1^0 \text{ (large loan account balance)} \\
\sum_j \lambda_j z_{2j} &\geq z_2^0 \text{ (small loan account balance);} \\
\sum_j \lambda_j z_{3j} &\geq z_3^0 \text{ (deposits account balance);} \\
\sum_j \lambda_j &= 1; y_1, y_2, y_3 \in Z_+; \lambda_j \geq 0 \text{ (} j = 1, 2, \dots, N \text{)}
\end{aligned} \tag{22}$$

In model (22) both the number of accounts as well as the average balance per account are incorporated as constraints, since both have implications for the labor requirement of a branch. The optimal numbers of these different types of accounts are required to be integers. However, the optimal amount of expenditure on any specific type of labor is not constrained to be an integer. Also, the optimal expenditure on a specific type of labor can exceed the actual.

Using this conceptual framework of cost minimization and the DEA model (22) above, we estimate the labor cost efficiency of the sample branches. The feasibility of these targets should be assessed by the branch management in light of any other constraints applicable to a specific branch.

It is worthwhile to argue here that cost minimization is the most appropriate economic objective of an Indian bank branch. While profitability is a meaningful objective at the bank level, it is not the appropriate model for a branch, especially in the Indian banking context. In a standard profitability model of branch banking (as proposed by McEachern and Paradi, 2007, Portela and Thanassoulis, 2005), the inputs considered are basically expenses towards interest outgo, remuneration and other operating cost. Outputs are revenues in the form of interest and non-interest earnings. Using the profitability model would imply that the deposit taking branches are loss making as interest expenses are much higher than interest earning. In the Indian context, branch profitability in practice, however, is not assessed by its interest income from loans or interest expenses towards deposits. Rather, as a corporate strategy, each commercial bank is considered to have an internal credit market, where a deposit oriented branch effectively sells its surplus funds to credit oriented branches at call money market rates that are aligned to

the repo rate. The overall economic viability of a branch is mainly judged based on its achievements over its targets that include imputed profitability along with customer service indicators.

5. Results

Chennai, Delhi, Kolkata and Mumbai are the largest metropolitan cities in the country and represent considerable divergence in terms of their social and economic environment. Mumbai is the capital of the state of Maharashtra in the western part of India. It is also the financial capital of the country where India's major stock exchange (BSE) is located and offers a thriving business environment. Not surprisingly, we observe that in 2014, our average branch in Mumbai has the highest balance of deposits (nearly ₹1.5 billion) and loans (about ₹0.5 billion). Chennai is the capital city of the state of Tamil Nadu in southern India which prides itself for its high skilled workers and superior work ethics. The balances of deposit and loan accounts for the average branch in Chennai were roughly ₹0.8 billion and ₹0.3 billion, respectively. Delhi, on the other hand is the capital of India and hence the most prominent city in terms of political and administrative power. Kolkata is the capital of the eastern state of West Bengal known for its inclination to art and culture.

Table 2 summarizes the labor cost efficiency of our sample of branches based on model (22). We measure the labor cost efficiency ($\bar{\gamma}_g$) of each branch first with respect to the metro-specific group frontier. Based on the respective metro frontiers, the average efficiency was 78.28% for our entire sample of branches in 2008 and improved to 81.71% in 2014. This implies that overall, in 2008 and 2014, the branches of this bank could reduce labor cost by about 22% and 18%, respectively, if they operated according to the best practices of branches in their respective metros. Comparison of the actual total cost (over all branches in a metro) with the sum of the minimum cost of each branch provides a measure of the average cost efficiency of branches in the metro. As shown in equations (19) and (20), the average metro-specific efficiency is equal to the cost-share weighted average efficiency of the branches within the metro. This is a more accurate measure than the simple mean of the efficiency of branches within a region as is often reported.¹⁷

¹⁷ The simple average efficiency of branches for any specific metro differed from the cost share weighted average measures reported in Table 2.

Table 2: Labor Cost Efficiency of Branches across Metros

	No. Branches	C ⁰	C*	C**	Cost Efficiency		
					Based on Group Frontier ($\bar{\gamma}_g$)	Based on Industry Frontier ($\bar{\bar{\gamma}}_g$)	Area Efficiency (μ_g)
2008							
Chennai	72	570,544	477,446	418,802	0.8368	0.7340	0.8772
Delhi	67	532,473	387,710	336,560	0.7281	0.6321	0.8681
Kolkata	123	1,154,641	855,790	605,604	0.7412	0.5245	0.7077
Mumbai	63	836,456	701,178	529,779	0.8383	0.6334	0.7556
All	325	3,094,115	2,422,124	1,890,745	0.7828	0.6111	0.7806
2014							
Chennai	72	810,950	653,159	581,381	0.8054	0.7169	0.8901
Delhi	67	866,234	721,990	568,745	0.8335	0.6566	0.7877
Kolkata	123	1,782,792	1,485,781	968,505	0.8334	0.5433	0.6518
Mumbai	63	1,394,808	1,105,714	757,102	0.7927	0.5428	0.6847
All	325	4,854,783	3,966,643	2,875,733	0.8171	0.5924	0.7250

Notes:

The nominal data for 2008 and 2014 are in current Rupees.

Cost figures C⁰, C*, C** are in ₹'000.

C⁰ = Sum of the actual cost of branches in a metro = $\sum C_j^0$;

C* = Sum of the minimum cost of branches in a metro based on the metro specific group frontier = $\sum C_j^*$;

C** = Sum of the minimum cost of branches in a metro based on the industry frontier = $\sum C_j^{**}$.

$\bar{\gamma}_g$ = Share weighted average cost efficiency of the metro relative to the group frontier. $\bar{\gamma}_g = \frac{\sum C_j^*}{\sum C_j^0}$

$\bar{\bar{\gamma}}_g$ = Share weighted average cost efficiency of the metro relative to the industry frontier. $\bar{\bar{\gamma}}_g = \frac{\sum C_j^{**}}{\sum C_j^0}$

$\mu_g = \frac{\bar{\bar{\gamma}}_g}{\bar{\gamma}_g} = \frac{\sum C_j^{**}}{\sum C_j^*}$.

An interesting question is how the branches in a specific metro perform with respect to the industry frontier constructed from branches across all the metros. Using equation (20a), we measure this cost efficiency ($\bar{\bar{\gamma}}_g$) of each branch and the summary results are reported in Table 2. We find that relative to the industry frontier Chennai branches have the highest efficiency in both years. Comparing across the two years we find that Delhi and Kolkata branches experienced a slight improvement in following the national best practices while the efficiency in the other two metro areas deteriorated, slightly for Chennai and considerably for Mumbai branches.

The average efficiency of branches measured against the metro specific frontier clearly reveals how efficiently the unit is managed. This information is particularly valuable to the branch managers. On the other hand, the average efficiency of branches in a metro measured against a common industry frontier reflects both the operational efficiency of the branches as well as differences in economic, institutional, and social conditions across the metros. As shown in equation (21), the ratio of the industry

efficiency and group efficiency offers a measure of the *area efficiency* of the metro. The area efficiency is an important measure in providing guidance to the regional and corporate level managers of the bank. The last column of Table 2 shows that Chennai has a high area efficiency of 87.72% in 2008 which improved to 89.01% in 2014. The area efficiency of the other three metros declined between 2008 and 2014. Kolkata and Mumbai have considerably low area efficiency in 2014.

Since our sample pertains to branches of the same bank, it is important to note that branches across the four metros have the same wage structure, employee training programs, as well as information technology systems to support the employees. They are also guided by the same corporate values and goals. However, these very same corporate level guidelines and incentives become more or less effective based on the overall environment of an individual metro. The major aspects of the environment include the (a) regulatory environment (b) social environment (c) local economic conditions (d) infrastructure and (e) industrial relations. Das et al. (2009) expounds on the cultural differences and value systems that have prevailed in the four metros. While the banking regulations are uniform across the states in India, heavy presence of state regulations in other aspects often influence the efficiency of branch performance across regions. Infrastructure differences can also play an important role. For example, wide power outages in a metro can lead to branches operating at less than their potential even with the same availability of computers and technology. Further, labor unions and strikes can also undermine the productive capability of branches.

Table 3: Frequency Distribution of Efficient Branches across Metros

	No. Branches	Based on Group Frontier		Based on Industry Frontier	
		No. of Efficient Branches	% of Efficient Branches	No. of Efficient Branches	% of Efficient Branches
2008					
Chennai	72	29	40.28%	13	18.06%
Delhi	67	22	32.84%	10	14.93%
Kolkata	123	32	26.02%	11	8.94%
Mumbai	63	19	30.16%	9	14.29%
All	325	102	31.38%	43	13.23%
2014					
Chennai	72	22	30.56%	13	18.06%
Delhi	67	28	41.79%	13	19.40%
Kolkata	123	37	30.08%	9	7.32%
Mumbai	63	22	34.92%	8	12.70%
All	325	109	33.54%	43	13.23%

Next we examine the distribution of branches performing on the cost frontier. From Table 3 we see that the percentage of branches on the metro specific group cost frontier was at a low of 26.02% and 30.08% for Kolkata in 2008 and 2014, respectively. Kolkata also has the lowest number of branches on the industry frontier in both the years. On the other hand, in 2008, Chennai had the highest percentage of branches on the group and industry frontiers with Delhi branches gaining the lead in 2014.

The results reported in Table 2 allow us to compare only the *means* of the actual sample. A more comprehensive test of equality of two distributions of efficiencies can be carried out by estimating the Kernel density functions through bootstrapping and performing the nonparametric Li Test (Li, 1996). We estimate the empirical density functions of the different efficiency measures across years and across groups and utilize the Li Test adapted by Simar and Zelenyuk (2006) for DEA. The results of these tests are presented in Table 4 which indicate that the cost efficiency distribution based on the industry frontier differs significantly between 2008 and 2014. Comparing the area efficiency distributions for each metro across the two years we find that the distributions do not differ significantly between 2008 and 2014, except for Kolkata branches. Lastly, we find from Table 4 that for 2008 the distribution of area efficiency for Chennai differs significantly only from the distribution for Kolkata. On the other hand, for 2014, the difference in area efficiency distributions between Chennai and each of the other metros is significant.

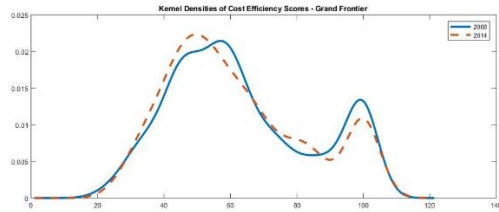
Table 4: Results from Tests of Equality of Efficiency Distribution

	Li-adapted Test Statistic	Bootstrap P- Value
Cost efficiency distribution based on the industry frontier: 2014 vs 2008	1.6815**	0.0190
Metro-specific Area Efficiency distribution: 2014 vs 2008		
Chennai	0.1018	0.9040
Delhi	-0.3376	0.6760
Kolkata	2.0573**	0.0190
Mumbai	-0.2667	0.7330
Metro-specific Area Efficiency distribution: Chennai vs Other Cities -2008		
Delhi	0.2579	0.7800
Kolkata	2.1280***	0.0000
Mumbai	-0.0069	0.9910
Metro-specific Area Efficiency distribution: Chennai vs Other Cities -2014		
Delhi	0.8958*	0.0560
Kolkata	2.212***	0.0000
Mumbai	0.8742*	0.0690

Note: *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively. The test statistic is constructed following the procedure in Simar and Zelenyuk (2006).

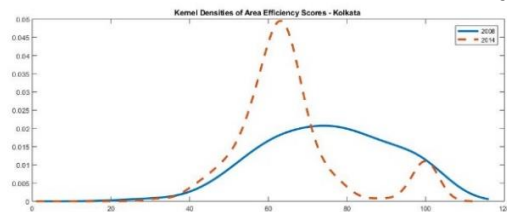
The relevant Kernel density plots for pairwise comparisons of different distributions are shown in Figures 2 through below for the cases where the Li Test statistic was significant.¹⁸

Figure 2: Kernel Densities of Cost Efficiency Scores of all Branches Based on the Industry Frontier



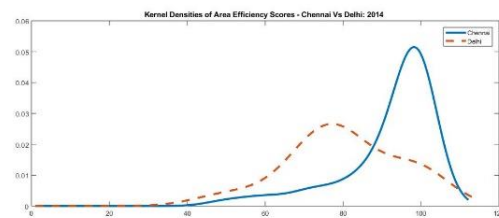
Note: 2008 - solid line. 2014 - dashed line

Figure 3: Kernel Densities of Area Efficiency Scores - Kolkata



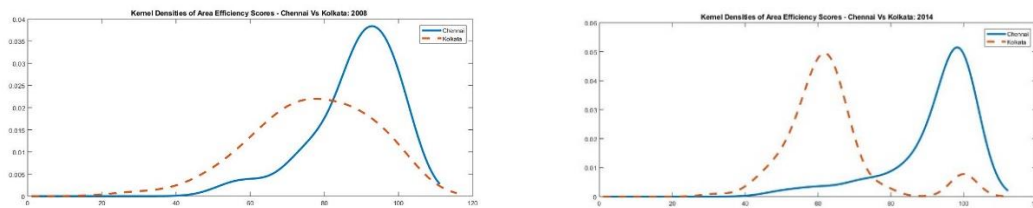
Note: 2008 - solid line. 2014 - dashed line

Figure 4: Kernel Densities Chennai versus Delhi 2014



Note: Chennai - solid line. Delhi - dashed line

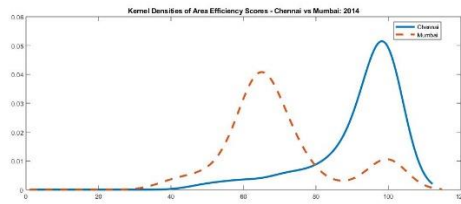
Figure 5: Kernel Densities Chennai versus Kolkata 2008 (Left Panel) and 2014 (Right Panel)



Note: Chennai - solid line. Kolkata - dashed line

¹⁸ The Kernel density plots for the other comparisons are available from the authors upon request.

Figure 6: Kernel Densities Chennai versus Mumbai 2014



Note: Chennai - solid line. Mumbai - dashed line

In the Indian banking sector, labor costs in the form of salaries and perks constitute 70% of operating costs and about 25% of total costs (Das et al., 2009). Managing labor costs efficiently at the branch level would have a significant impact on the overall operating efficiency of the bank. Our analysis reveals that if all 325 branches were able to eliminate their labor cost inefficiency and performed based on the best practices of branches within the metro in which they are located, then the bank in our study would be able to cut its labor costs by roughly ₹888 million from ₹4.855 billion to ₹3.967 billion in 2014 (a labor cost saving of 18.29%). If the bank was able to incentivize all its 325 branches to follow the best practices of their national role models, then the bank would be able to cut its labor costs by ₹1.979 billion from ₹4.855 billion to ₹2.875 billion in 2014.¹⁹

Comparison of the actual and optimal total labor costs fail to reveal the potential savings in labor cost by employee categories. Tables 5A and 5B shed light into this aspect of our analysis. We find that in 2014, across the four metros, the branches in our sample are over spending on officers by 23%, on clerks by 20%, and on subordinates by 29% compared to the optimal cost based on metro specific frontiers. Only in case of clerks, there was some improvement in the utilization rate compared to 2008. It should be noted that cost minimization does not require that all categories of labor be reduced. In fact, the branches could be under- utilizing some categories of labor compared to the benchmark branch. Table 6A and 6B presents the number of branches with over, under, and optimal utilization of different categories of labor in 2008 and 2014. In 2014, based on the group frontiers, we find that out of the 325 branches in the sample, 181 are spending over spending, 34 are under spending, and the remaining 110 are spending the optimal amount on officers. The comparable frequencies for clerks during the same year are 194 over spending, 22 under spending, and 109 are spending the optimal amount. Finally, for subordinates, the frequencies are 174 over spending, 39 under spending and the remaining 112 are spending the optimal amount. The comparable figures for the different categories of labor for 2008 can be found in the upper panel of Table 6A. A comparison between the two years shows that the number of branches spending too

¹⁹ The savings potential with respect to each metro of course depends on the measured inefficiency at the branches as well as the size and number of branches.

much on clerks has declined from 206 in 2008 to 194 in 2014. This is reflected in the decline in the rate of overutilization for clerks from 32% in 2008 to 20% in 2014 seen in Tables 5A and 5B.

Table 5A: Utilization Rates for Different Categories of Labor (2008)

No. Branches	Chennai 72	Delhi 67	Kolkata 123	Mumbai 63	All 325
Officers					
c_1^0	202,373	209,843	366,716	253,985	1,032,917
c_1^*	173,399	165,246	286,460	226,034	851,139
c_1^{**}	162,528	141,783	226,733	191,655	722,699
η_1^*	1.17	1.27	1.28	1.12	1.21
η_1^{**}	1.25	1.48	1.62	1.33	1.43
Clerks					
c_2^0	263,930	229,942	566,011	422,080	1,481,963
c_2^*	221,154	152,016	404,866	346,892	1,124,928
c_2^{**}	177,467	131,039	254,604	241,282	804,392
η_2^*	1.19	1.51	1.40	1.22	1.32
η_2^{**}	1.49	1.75	2.22	1.75	1.84
Subordinates					
c_3^0	104,241	92,688	221,914	160,391	579,234
c_3^*	82,892	70,448	164,464	128,252	446,057
c_3^{**}	78,807	63,739	124,267	96,842	363,655
η_3^*	1.26	1.32	1.35	1.25	1.30
η_3^{**}	1.32	1.45	1.79	1.66	1.59

Notes:

The nominal data for 2008 and 2014 are in current Rupees. See note below Table 1.

Cost figures c^0 , c^* , c^{**} are in ₹'000.

c_i^0 = Sum of actual cost of branches in a metro on the individual component of labor = $\sum_j c_{ij}^0$;

c_i^* = Sum of the minimum cost of branches in a metro on the individual component of labor (based on the metro specific group frontier) = $\sum_j c_{ij}^*$;

c_i^{**} = Sum of the minimum cost of branches in a metro on the individual component of labor (based on the industry frontier) = $\sum_j c_{ij}^{**}$;

$\eta_i^* = \frac{c_i^0}{c_i^*}$,

$\eta_i^{**} = \frac{c_i^0}{c_i^{**}}$.

Table 5B: Utilization Rates for Different Categories of Labor (2014)

No. Branches	Chennai 72	Delhi 67	Kolkata 123	Mumbai 63	All 325
Officers					
c_1^0	366,807	462,495	861,198	631,280	2,321,780
c_1^*	294,113	382,080	714,102	501,531	1,891,827
c_1^{**}	276,674	285,975	470,727	363,797	1,397,173
η_1^*	1.25	1.21	1.21	1.26	1.23
η_1^{**}	1.33	1.62	1.83	1.74	1.66
Clerks					
c_2^0	343,897	303,947	686,175	579,820	1,913,839
c_2^*	281,828	261,895	581,182	470,626	1,595,531
c_2^{**}	243,428	228,380	393,443	310,477	1,175,728
η_2^*	1.22	1.16	1.18	1.23	1.20
η_2^{**}	1.41	1.33	1.74	1.87	1.63
Subordinates					
c_3^0	100,246	99,792	235,418	183,708	619,164
c_3^*	77,217	78,015	190,497	133,556	479,285
c_3^{**}	61,279	54,389	104,335	82,827	302,831
η_3^*	1.30	1.28	1.24	1.38	1.29
η_3^{**}	1.64	1.83	2.26	2.22	2.04

Notes: See Table 5A.

Table 6A: Branches with Over, Under, and Optimal Utilization of Different Categories of Labor (2008)

<u>Based on Group Frontier</u>		Officers			Clerks			Subordinates		
		Number of Branches with								
No. Branches		$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$
Chennai	72	30	30	12	39	29	4	35	29	8
Delhi	67	43	22	2	42	22	3	41	22	4
Kolkata	123	78	32	13	86	32	5	80	32	11
Mumbai	63	33	19	11	39	19	5	38	19	6
All	325	184	103	38	206	102	17	194	102	29

<u>Based on Industry Frontier</u>		Officers			Clerks			Subordinates		
		Number of Branches with								
No. Branches		$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$
Chennai	72	51	13	8	56	13	3	43	13	16
Delhi	67	56	10	1	56	10	1	47	10	10
Kolkata	123	109	11	3	111	11	1	99	11	13
Mumbai	63	48	9	6	53	9	1	48	9	6
All	325	264	43	18	276	43	6	237	43	45

Table 6B: Branches with Over, Under, and Optimal Utilization of Different Categories of Labor (2014)

<u>Based on Group Frontier</u>		Officers			Clerks			Subordinates		
		Number of Branches with								
No. Branches		$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$
Chennai	72	36	22	14	48	22	2	36	25	11
Delhi	67	38	28	1	34	28	5	33	28	6
Kolkata	123	68	38	17	74	37	12	71	37	15
Mumbai	63	39	22	2	38	22	3	34	22	7
All	325	181	110	34	194	109	22	174	112	39

<u>Based on Industry Frontier</u>		Officers			Clerks			Subordinates		
		Number of Branches with								
No. Branches		$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$	$\eta > 1$	$\eta = 1$	$\eta < 1$
Chennai	72	44	13	15	57	13	2	42	13	17
Delhi	67	54	13	0	46	13	8	48	13	6
Kolkata	123	113	9	1	114	9	0	104	11	8
Mumbai	63	54	8	1	54	8	1	50	9	4
All	325	265	43	17	271	43	11	244	46	35

A general observation across our models and different construction of frontiers is that bank branches in India exhibit significant inefficiencies in labor use. This result is in agreement with the findings of Das et al. (2009) and Ray (2016) on Indian bank branch efficiency. It is interesting to compare these results for Indian bank branches with that of branches of other countries surveyed by Paradi and Zhu (2013). They report that most branch level studies reveal high efficiency level of branches. One reason for the measured high efficiency levels in many of the other studies (especially in the context of other developing countries) could be the very small data sets used, which reduces the discretionary power of the analysis. Another reason could be that the branch management in western countries enjoy much greater flexibility in scheduling the number and hours of mostly part time employees. In contrast, bank employees in India are full time and permanent. Hence, adjusting the work force to business conditions is far more difficult.

We also find that the cost efficiency distributions (based on the industry frontier) between the two years differ significantly, with a slightly lower average efficiency in 2014. In any year, the industry frontier is supported by the efficient branches. In a period of technological regime change, a limited number of innovative branches may be able to reorganize their internal organization appropriately to take

advantage of computerization while a majority of the branches lag behind. This would result in an overall lower measured efficiency of a large number of branches. Looking at the measured labor utilization rates for clerical employees (as reported in Tables 5A and 5B), we find that there has been a significant improvement in 2014. This appears to be a direct consequence of computerization of manual tasks performed by these type of employees. Specifically, traditional jobs like passbook updating, cash deposit, verification of know-your-customer details, salary uploads which were primarily by clerks even in 2008 have mostly gone digital by 2014 and there has been redeployment of employees to more productive functions. At the same time there has been increasing demands on officers to acquire new analytical skills. To the extent that officers at many branches are still in the learning mode, the measured labor utilization rate of officers show a deterioration in 2014.

6. Conclusion

In this study, we address the issue of operational efficiency of Indian bank branches and examine the efficiency of 325 branches of a major Indian public sector bank across four major metropolitan cities, Chennai, Delhi, Kolkata, and Mumbai in 2008 and 2014.

Our results indicate that there is significant labor cost inefficiency in the operations of the branches. Overall Chennai branches are much closer to their metro specific best-practice frontier compared to branches of other metros from their respective metro-specific frontier. The low area-efficiency of Kolkata branches is of particular concern; even branches that perform at the frontier of their metro specific technology are generally far behind in terms of the industry best practices. Additional measures need to be taken by the bank at the corporate and regional levels in offsetting the negative effects of poor area efficiency in Kolkata and Mumbai branches.

As competitive pressures in the Indian banking sector intensify, corporate level management must focus on adopting best practices from external sources (other banks and outside experts) as well as best practice transfer within units (branches) of its own organization. Our study reveals that for this public sector bank, efficient branches within the same metro should be identified to share local best practices and efficient branches in Chennai could serve as ‘sources’ of best practice transfer of national best practices. Kolkata branches would indeed benefit as ‘recipients’ of such knowledge transfer. Of course this process should take into account local idiosyncrasies. As recognized in the management literature, the ability to successfully transfer best practices internally may be difficult due to motivational factors or due to other sources of stickiness such as the recipients’ lack of absorptive capacity, causal ambiguity in identifying why the ‘source’ is efficient, or arduous relationship between the ‘source’ and the ‘recipient’ (Szulanski, 1996) . An important contribution of this study is to identify the branches that are performing poorly compared to their peers in the same metro area as prime candidates for review. The interregional

differences, while beyond the control of the local managers, may be relevant for corporate planners as they chart out the growth strategy and location decisions in an increasingly globalized economy.

We end with a special note about the proliferation of ATMs, Debit/Credit Cards and online banking in 2014. It should be noted that we do not include data on ATMs, Debit/Credit Cards and electronic transactions into our model since such data are not available at the branch level. Further, it should be noted that with the advent of a core banking platform, ATMs are not dedicated to a branch and one can carry out transactions relating to accounts located in any branch of the bank from ATMs (or online) from anywhere in the country. There is no reason, therefore, to presume that ATMs will affect performance across branches unevenly. While it will affect the efficiency of clerical staffs and officers, it should not introduce any heterogeneity across branches.

References

- Athanassopoulos, A. D. (1997) Service Quality and Operating Efficiency Synergies for Management Control in the Provision of Financial Services: Evidence from Greek Bank Branches, *European Journal of Operational Research*, 98, 300 -313.
- Badunenko, O., Kumbhakar, S. C. (2017) Economies of Scale, Technical Change and Persistent and Time Varying Cost Efficiency in Indian Banking: Do Ownership, Regulation and Heterogeneity Matter? *European Journal of Operational Research*, 260, 789-803.
- Banker, R. D., Chan, H., Natarajan, R. (2007) Estimating DEA Technical and Allocative Inefficiency Using Aggregate Cost or Revenue Data, *Journal of Productivity Analysis*, 27, 115 – 121.
- Battese, G. E., Rao, D. S. P. (2002) Technology Gap, Efficiency, and a Stochastic Metafrontier Function, *International Journal of Business and Economics*, 1, 87-93.
- Battese, G. E., Rao, D. S. P., O'Donnell C. J. (2004) A Metafrontier Production Function for Estimation of Technical Efficiencies and Technology Gaps for Firms Operating Under Different Technologies, *Journal of Productivity Analysis*, 21, 91-103.
- Benston, G. J. (1965) Branch Banking and Economies of Scale, *Journal of Finance*, 20, 312-331.
- Berger, A. N. (2007) International Comparisons of Bank Efficiency, *Financial Markets, Institutions and Instruments*, 16, 119-44.
- Berger, A. N., Humphrey, D. B., (1997) Efficiency of Financial Institutions: International Survey and Directions for Future Research, *European Journal of Operational Research*, 98, 175-212.
- Berger, A. N., Leusner, J.H., Mingo, J. J. (1997) The Efficiency of Bank Branches, *Journal of Monetary Economics*, 40, 141-162.
- Bhandari, A., Ray, S. C. (2011) Technical Efficiency in the Indian Textile Industry: A Non-parametric Analysis of Firm-Level Data, *Bulletin of Economic Research*, 64, 109-124.
- Bhaumik, S. K., Dimova, R. D., (2004) How Important is Ownership in a Market with Level Playing Field? The Indian Banking Sector Revisited, *Journal of Comparative Economics*, 32, 165-180.
- Camanho, A. S., Dyson, R. G. (1999) Efficiency, Size, Benchmarks and Targets for Bank Branches: An Application of Data Envelopment Analysis, *The Journal of the Operational Research Society*, 50, 9, 903-915.
- Casu, B., Ferrari, A., Zhao, T. (2013) Regulatory Reform and Productivity Change in Indian Banking, *The Review of Economics and Statistics*, 95, 3, 1066-1077.
- Das, A., Ghosh, S. (2006) Financial Deregulation and Efficiency: An Empirical Analysis of Indian Banks during the Post-Reform Period, *Review of Financial Economics*, 15, 193-221.

- Das, A., Ray, S. C., Nag, A. (2009) Labor-Use Efficiency in Indian Banking: A Branch-Level Analysis, *Omega*, 37, 411-425.
- Drekker, D., Post, T. (2001) A Quasi-Concave DEA Model With an Application for Bank Branch Performance Evaluation, *European Journal of Operational Research*, 132, 296-311.
- Färe, R., Grosskopf, S., Lovell, C. A. K. (1994) *Production Frontiers*, Cambridge University Press, Cambridge, England.
- Fethi, M. D., Pasiouras, F. (2010) Assessing Bank Efficiency and Performance with Operational Research and Artificial Intelligence Techniques: A Survey, *European Journal of Operational Research*, 204, 189-98.
- Fujii, H., Managi, S., Matousek, R. (2014) Indian Bank Efficiency and Productivity Changes with Undesirable Outputs: A Disaggregated Approach, *Journal of Banking and Finance* 38, 41-50.
- Hayami, Y. and Ruttan, V. W. (1971) *Agricultural Development: An International Perspective*, Baltimore, Johns Hopkins University Press.
- Kuosmanen, T., Matin, R. K. (2009) Theory of Integer-Valued Data Envelopment Analysis, *European Journal of Operational Research*, 192, 2, 658-667.
- Li, Qi (1996) Nonparametric Testing of Closeness Between Two Unknown Distribution Functions, *Econometric Reviews*, 15, 261-274.
- Matin, R. K., Kuosmanen, T. (2009) Theory of Integer-Valued Data Envelopment Analysis under Alternative Returns to Scale Axioms, *Omega*, 37, 5, 988-995.
- McAllister, P. H., McManus, D. (1993) Resolving the Scale Efficiency Puzzle in Banking, *Journal of Banking and Finance*, 17, 389-405.
- McEachern, D., Paradi, J. C. (2007) Intra- and Inter-Country Bank Branch Assessment Using DEA, *Journal of Productivity Analysis*, 27, 123-136.
- Paradi, J. C., Vela, S. A., Zhu, H., (2010) Adjusting For Cultural Differences, A New DEA Model Applied to a Merged Bank, *Journal of Productivity Analysis*, 33, 109-123.
- Paradi, J. C., Zhu, H. (2013) A Survey of Bank Branch Efficiency and Performance Research with Data Envelopment Analysis, *Omega*, 41, 61-79.
- Portela, M.C.A.S., Thanassoulis, E. (2005) Profitability of a Sample of Bank Branches and its Decomposition into Technical and Allocative Components, *European Journal of Operational Research*, 162, 850-866.
- Quaranta, A. G., Raffoni, A., Visani, F. (2018) A Multidimensional Approach to Measuring Bank Branch Efficiency, *European Journal of Operational Research*, 266, 746-760.
- Ram Mohan, T. T., Ray, S. C. (2004) Comparing Performance of Public and Private Sector Banks: A Revenue Maximization Efficiency Approach, *Economic and Political Weekly*, 39, 1271-1275.

- Ray, S. (2016) Cost Efficiency in an Indian Bank Branch Network: A Centralized Resource Allocation Model, *Omega*, 65, 69-81.
- Ray, S. C., Das, A. (2010) Distribution of Cost and Profit Efficiency: Evidence from Indian Banking, *European Journal of Operational Research*, 201, 297-307.
- Ray, S. C., Mukherjee, K. (2017) Data Envelopment Analysis with Aggregated Inputs and a Test of Allocative Efficiency when Input Prices Vary Across Firms, *Data Envelopment Analysis Journal*, 2, 141-161.
- Sathye, M. (2003) Efficiency of Banks in a Developing Economy: The Case of India, *European Journal of Operational Research*, 148, 662-671.
- Saha, A. and Ravisankar, T. S. (2000) Rating of Indian Commercial Banks: A DEA Approach, *European Journal of Operational Research*, 124, 187-203.
- Sensarma, R. (2006) Are Foreign Banks Always the Best? Comparison of State-Owned, Private and Foreign Banks in India, *Economic Modelling*, 23, 717-735.
- Sherman, H. D., Gold, F. (1985) Bank Branch Operating Efficiency: Evaluation with Data Envelopment Analysis, *Journal of Banking and Finance*, 9, 297-315.
- Simar, L., Zelenyuk, V. (2006) On Testing Equality of Distributions of Technical Efficiency Scores, *Econometric Reviews*, 25, 497-522.
- Szulansky, G. (1996) Exploring Internal Stickiness: Impediments to the Transfer of Best Practice within the Firm, *Strategic Management Journal*, 17, 27-43.
- Tzeremes, N. G. (2015) Efficiency Dynamics in Indian Banking: A Conditional Directional Distance Approach, *European Journal of Operational Research*, 240, 807-818.
- Wu, D., Yang, Z., Liang, L. (2006) Efficiency Analysis of Cross-Region Bank Branches Using Fuzzy Data Envelopment Analysis, *Applied Mathematics and Computation*, 186, 1600-1611.
- Zhao, T., Casu, B., Ferrari, A. (2010) The Impact of Regulatory Reforms on Cost Structure, Ownership, and Competition in Indian Banking, *Journal of Banking and Finance*, 34, 246-254.