Data Envelopment Analysis: A Tool of Measuring Efficiency in Banking Sector

Filzah Mohamed Othman¹, Nor Aiza Mohd-Zamil²*, Siti Zaleha Abdul Rasid³, Amin Vakilbashi⁴, Mozhdeh Mokhber⁵

¹International Business School, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, ²International Business School, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, ³International Business School, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, ⁴International Business School, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, ⁵International Business School, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia, *Email: noraiza@ibs.utm.my

ABSTRACT

The present paper examined the review of literature related to measuring relative efficiency of banks using data envelopment analysis (DEA). The efficiency of banks is measure through the ability of the individual bank to maximise output given a certain level of input. By measuring its efficiency, it can serves as early warning or benchmark of its performance and it can define future improvement in various area such as managerial, technology or socio-economic. DEA is comprises of two basic model that are DEA Charnes-Cooper-Rhodes model with constant return to scale assumption and DEA Banker-Charnes-Cooper model with variable return to scale assumption. In banking industry, DEA is using two approaches that are production or intermediation approach. The former highlights banks as delivering services in the form of transaction and the later assumes banks intermediate funds between surplus units to deficit unit. The study of efficiency in banks with similar economic and political condition is important as banks operate in parallel.

Keywords: Data Envelopment Analysis, Efficiency, Banking

JEL Classifications: G2, M2, M4

1. INTRODUCTION

The foundation of productivity in service industry, specifically in banking sector, generally is measured based on two key concepts, namely effectiveness and efficiency (Sherman and Zhu, 2006). Effectiveness is referring to the ability of the bank to set and achieve its goals and objectives, while efficiency refers to ability of the bank to produce output with minimal resources or input, or commonly defined as the ratio of outputs over inputs (Sherman and Zhu, 2006; Chen et al, 2008).

Thus, many literatures use the terms productivity and efficiency interchangeably. The efficiency of financial institutions has been widely and extensively studied in the last few decades. For financial institutions, efficiency implies improved profitability, greater amount of funds channeled in, better prices and service quality for consumers and greater safety in terms of improved capital buffer in absorbing risk (Berger et al., 1993). Although banks main focuses are to find ways to generate new funds and lending funds at higher rate, they have developed concerns in managing their operational productivity in order to ensure higher profitability and consistently attract more investors (Sherman and Zhu, 2006).

Basically, efficiency can be defined as the ratio of output to input; and more output per unit of input indicates greater efficiency while maximum output per unit of input reflects optimum efficiency (Cooper et al., 2006; Sherman and Zhu, 2006). Efficiency measurement determine how firm can maximize its output and profit and at the same time minimize its cost (Mokhtar et al., 2008). The importance of efficiency measurement is to enable managers to benchmark bank performance and define areas of inefficiency for future improvements (Mostafa, 2007). The areas of inefficiency is not limited to the result of poor management performance alone, instead it might be due to managerial, technological and socio-economic (Sherman and Zhu, 2006).
2. LITERATURE REVIEW

2.1. Efficiency Classifications
According to Sherman and Zhu (2006), overall productivity of a bank depends on four components of efficiency classification as shown in Figure 1 and they are:
1. Technical efficiency: Also known as global efficiency measures the ability of banks to produce actual outputs with fewer inputs, or less resources used indicates higher efficiency;
2. Scale efficiency: Refers to the optimal activity volume level whereby inefficiency may arise if goods or services are produced above or below optimal level that resulted in added fixed cost;
3. Price efficiency: Bank could increase its efficiency if it could purchase the inputs (human capital and material) at lower price without sacrificing the quality;
4. Allocative efficiency: Measure the optimal mix of several inputs in order to produce products or services, such as banks incorporate automatic teller machines (ATM) and Internet banking for capital labour tradeoffs to increase efficiency (Sherman and Zhu, 2006).

In addition, by definition, technical efficiency refers to the firm ability to maximize output with the given inputs or; produce same level of outputs with minimization of inputs; while allocative efficiency refers to the optimum arrangement of inputs and output at a specific price (Cooper et al., 2006). Technical inefficiency may arise in the conditions where banks produce more outputs with the actual inputs or when bank produce actual output with fewer inputs (Sherman and Zhu, 2006), or generally speaking technical inefficiency exists when banks are wasting some of inputs (Mester, 2003).

Besides the general classification above, efficiency can also be categorized into X-efficiency. X-efficiency measure how productive a bank uses its input to create outputs from the aspect of selecting the appropriate inputs. Conceptually, the measurement of X-efficiency can further be broken down into two components of efficiency for extended efficiency analysis which are cost efficiency and profit efficiency base on economic concept of cost-minimization and profit-maximization (Mester, 2003; Mensah, 2012).

Profit efficiency takes into account the effects of cost and revenues; as it measures the ratio of actual bank profit to maximum level of profit that achievable by most efficient bank. The more complicated efficiency analysis for banks normally involves incorporating risk-return trade-offs (Mester, 2003). On this note, the selection of variables for input-output relationships and the efficiency model employed that determined the type of efficiency under investigation (Mostafa, 2011).

2.2. Data Envelopment Analysis (DEA) and its Basic Models
DEA was first developed by Farrel in 1957, which later been modified by Charnes-Cooper-and Rhodes (CCR) in 1978 (Klimberg et al., 2009). It is a non-parametric method that utilizes linear programming to measure the level of efficiency of comparable decision-making units (DMU) by employing multiple inputs and outputs (Klimberg et al., 2009). This technique of measuring efficiency was first introduced by Farrel in 1957 based on the basic theory of production on single input and single output such as “output per work hour” in a form of ratio (Ayadi et al., 1998; Cooper et al., 2006; Sherman and Zhu, 2006).

\[
\text{Efficiency} = \frac{\text{Output}}{\text{Input}}
\]  

However, this measurement does not entirely represent efficiency as commonly multiple inputs are used to produce single or more outputs, which lead to the modification of original equation to include measurement of multiple inputs and multiple outputs (Zhu and Sherman, 2006). This concept was further extended into basic CCR DEA model developed by CCR in 1978 by altering the original equation to (Ayadi et al., 1998; Zhu and Sherman, 2006; Cooper et al., 2006).

\[
\text{Efficiency} = \frac{\text{Weighted sum of output}}{\text{Weighted sum of input}}
\]  

In DEA, methods to measure efficiency of DMUs are referred to a group of firms under study such as banks, hospital etc. DEA is a most accurate technique to measure efficiency given limited number of DMUs (i.e., banks) (Cooper et al., 2006; Klimberg et al., 2009; Hassan et al., 2009; Ahmad and Luo, 2010).

The DEA model was first modified by Sherman to measure banks performance in 1984, and since then, was extensively used by banking industry around the world to measure banks operational efficiency (Sherman and Zhu, 2006). DEA allows measurement of efficiency from multiple inputs and multiple outputs within multiple DMUs (Sherman and Zhu, 2006).

Accordingly, the mathematical equation to find the maximum efficiency of DMUs using weighted input-output efficiency measure can be expressed as Model 1 (Cooper et al., 2006; Sherman and Zhu, 2006; Ramanathan, 2007; Chen et al., 2008):

\[
\text{max} \sum_{j=1}^{J} v_{mj}y_{mj} \quad \text{subject to} \quad \sum_{i=1}^{I} u_{mi}x_{mi} = 1; n=1, 2,..., N \nonumber \\
0 \leq v_{mj}, u_{mi} \leq 1; i=1, 2,..., I; J=1, 2,..., J
\]  

Figure 1: General classification of bank efficiency (Adapted from Sherman and Zhu, 2006)
Where:

\( N \): Total number of DMUs

\( J \): Weighted sum of outputs

\( I \): Weighted sum of inputs

\( M \): The base DMU (calculating \( m \)th DMU)

\( N \): DMUs

\( I \): Inputs

\( J \): Outputs

\( v_{mj} \): Weights for output

\( u_{mi} \): Weights for input.

Since the above equation is in the fractional function, it is difficult to compute, thus, CCR (1978) transform the equation into linear programming equation by setting the denominator of the ratio to one or unity to form a linear programming equation Model 2 or equally known as output-maximization CCR model (Cooper et al., 2006; Sherman and Zhu, 2006; Ramanathan, 2007; Chen et al., 2008):

\[
\begin{align*}
\max & \sum_{j=1}^{J} v_{mj} y_{mj} \\
\text{Such that} & \\
\sum_{i=1}^{I} u_{mi} x_{mi} & = 1; \\
\sum_{j=1}^{J} v_{mj} y_{mj} - \sum_{i=1}^{I} u_{mi} x_{mi} & \leq 0; \quad n = 1, 2, \ldots, N \\
v_{mj}, u_{mi} & \geq 0; \quad i = 1, 2, \ldots, I; \quad j = 1, 2, \ldots, J
\end{align*}
\]

(Modl 2)

When DEA is employed to measure banks efficiency for a set of DMUs, the linear programming algorithm will calculate the efficiency of each DMU given the identical inputs and outputs variables to find the maximum ratio of weighted sum of output to the weighted sum of input (most efficient DMU) and to be used as benchmark against other DMUs, causing the best-practice DMUs to lie on the efficient frontier line. It means the best-practice units are relatively efficient and identified by DEA efficiency score as 100% (efficiency = 1).

Charnes et al. (1979) imposed non-negativity restrictions to ensure inputs and outputs have positive weight values, so as the efficiency score assigned will be between 1 and 0, and no efficiency index greater than one. The less productive units or inefficiency are identified with efficiency score of <100% (efficiency <1). The relative units to this frontier represent the degree of inefficiency. Graphically, the Figure 2 below illustrates the production frontier of the CCR Model, where it calculates most efficient DMUs on diagonal line across the area where frontier and other DMUs lies (production possibility sets).

The above explanations were derived from Soteriou and Stavrinides (1997), Cooper et al. (2006), Sherman and Zhu (2006), Ramanathan (2007), Chen et al. (2008), Hassan et al. (2009), Klimberg et al. (2009), and Yahya et al. (2012).

\[
\begin{align*}
\sum_{n=1}^{N} y_{nj} \lambda_n & \geq y_{mj}; \quad j = 1, 2, \ldots, J \\
\sum_{n=1}^{N} x_{ni} \lambda_n & \leq \theta_{mi} x_{mi}; \quad i = 1, 2, \ldots, I \\
\lambda_n & \geq 0; \quad n = 1, 2, \ldots, N; \quad \theta_{mi} \text{ free}
\end{align*}
\]

Where, \( \Theta_m = \text{Efficiency ratio of } m \text{th DMU.} \)

In input-minimization DEA CCR model which also referred to dual model in DEA literatures, the DMU is comparatively efficient if and only if the optimal values of its efficiency ratio (score), \( \Theta_m \) is equals one or unity (Cooper et al., 2006; Ramanathan, 2007; Wu and Wu, 2010).

In an attempt to use both input-oriented and output-oriented models to calculate DEA efficiency score for 55 banks in the Gulf Cooperation Council (GCC), Ramanathan (2007) discovered that both models generated similar results. This suggested that there is no obvious difference in efficiency score generated by both models. Thus, no misleading interpretations of DEA score if either one model is chosen.

According to Sherman and Zhu (2006), the basic CCR model developed by CCR assuming constant return to scale (CRS).
Efficient frontier in DEA can be derived using five alternative of return to scale assumption, which each describes the rate of substitution between inputs and outputs either to be increasing, constant, or decreasing within each segment of the frontier (Sherman and Zhu, 2006). The five alternatives are: (1) Increasing return to scale: A condition when there is proportionate increase of output and input causing DMU to be inside the frontier; (2) CRS: A condition when there is proportionate increase or decrease of input or output causing the DMU to be moved along the frontier line or above it, and provide meaningful measurement of technical and scale efficiency without having data on input price or cost; (3) variable return to scale (VRS): Is used when CRS assumption is not satisfied or there is no economies of scale, and efficiency of DMU on efficient frontier is interpreted as pure technically efficient; (4) non-increasing return to scale: A condition of DMU not being on the frontier line; and (5) non-decreasing return to scale: A condition of DMU not being on the frontier line (Ong et al, 2003; Sherman and Zhu, 2006; Ramanathan, 2007; Tahir and Haron, 2008).

According to Charnes et al. (1994), when DEA efficiency scores were calculated using both CRS and VRS model, CRS efficiency scores are less than or equal to the corresponding VRS efficiency score, due to the difference scale size of each DMUs. Ramanathan (2007) attempted to validate this finding by measuring efficiency of 55 banks in GCC using DEA with both assumptions, CRS and VRS on same data sets, and discovered that the results is consistent with the above, showing average CRS efficiency scores are less than average VRS efficiency scores for each country in GCC.

2.2.2. DEA Banker-Charnes-Cooper (BCC) model
The first extension of basic CCR model is called the DEA BCC model developed by BCC in 1984, with other criteria are the same as CCR except it complement the equation to measure input excesses and output shortfalls (Cooper et al., 2006; Ong et al., 2003). BCC model includes convexity condition with non-negative element constraints. The DEA BCC model equation, Model 4 (Cooper et al., 2006; Chen et al, 2008):

\[
\max_{\bar{u}, \bar{v}} \sum_{j=1}^{q} u_j Y_{kj} - u_0
\]

Subject to,

\[
\sum_{j=1}^{q} u_j Y_{kj} - u_0 \leq 0
\]

\[
\sum_{i=1}^{p} v_i X_{ki} \leq 1
\]

Where,

- \(P\): Inputs
- \(u_j\): Weights of output (virtual value)
- \(v_i\): Weights of input (virtual value)
- \(u_o\): Scalar free in sign (positive or negative or 0).

Basically, in BCC model, the formula calculates the efficiency of DMUs and most efficient DMUs that lie on the convex line creating efficient frontier after passing through the area of DMUs (production possibility set). The Figure 3 graphically illustrates production frontier of BCC model.

2.3. Comparison between DEA BCC Model and CCR Model
According to Cooper et al. (2006), CCR model calculate the proportional efficiency, but does not measure the input excess and output shortfalls, which complemented by BCC model. For further comparisons between BCC Model and CCR Model, the Figure 4 is referred.

Adapted from Cooper et al. (2006), differences between the two models are explained. Considering they are four DMUs, denoted...
as point A, B, C, D that utilize same type of input to produce same type of output, the CCR efficient frontier in the red dotted line that intersect with B from origin; while the BCC efficient frontier is the black bold line that touched A, B, C. Therefore, only B is CCR-efficient, while A, B, C are BCC-efficient. Thus, more efficient DMUs can be generated through BCC model and this is consistent with Mester (2003) and Burger and Moormann (2008) that discovered that BCC model are more relax allowing more DMUs on efficient frontier line.

Further, given the DMU D, BCC-efficiency can be calculated through ratio of PR/PD; while the CCR-efficiency value is determined through ratio of PQ/PD. The results show that the value of CCR-efficiency is smaller and never exceed BCC-efficiency value. With this, Cooper et al. (2006) theory is supported by Sherman and Zhu (2006).

Inefficiency for DMU D, is the distance from D to S, calculated through BCC model as ST/DT. While inefficiency of DMU D for CCR model is the distance from D to the point that intersect with CCR-efficiency line (dotted red line) from ST. Therefore, given the same level of input, DMU D can achieve optimum efficiency by increasing SD amount of output (output-oriented) (Cooper et al., 2006).

In contrary, the inefficiency for DMU D using CCR model can simply be identified by 1/(PQ/PD) creating a reciprocal relation between input-output (Sherman and Zhu, 2006). Thus, in order for DMU D to achieve optimal efficiency using CCR model, input should be reduced to PQ while producing same level of output (input-oriented) (Cooper et al., 2006; Sherman and Zhu, 2006).

Some other DEA models which were used in more advanced and detail research are slack-based model, hybrid model, Russell model and quality-adjusted (Q-DEA) model.

2.4. DEA as Efficiency Measurement Tool
The measurement of bank efficiency consists of two approaches, namely production and intermediation. Production approach emphasized banks as firm delivering services in the form of transactions, while in intermediation approach, banks assumed intermediating funds between savers and investors (Mostafa, 2011). Production approach assumed banks as the producers of loan and deposit for borrowers and depositors using traditional factor of production; capital, land, and labour (Taufiq et al., 2009). Intermediation approach is more appropriate in evaluating banking sectors as bank collect funds and transforms them into loans and other earning assets (Mokhtar et al., 2008). Thus, all literatures on measuring banks efficiency quoted in this literature employed either intermediation approach or production approach.

As financial intermediary, banks main function is to facilitate the distribution of financial resources effectively and efficiently in a form of investments and received financial benefits in return (Batchelor, 2005). Simply put, banks borrow funds from depositors to lend to others; hence deposits are considered as inputs (Mester, 2003; Chen et al., 2008). Figure 5 depicts the intermediation function for banks.

This intermediation approach applicable to both bank streams, conventional and Islamic banking and they are illustrated in the Figures 6 and 7 respectively.

Intermediation approaches in Islamic banks practice no exchange of interest between borrowers or depositors with the bank; rather the profit allocation is based on profit-sharing agreements between them (Sufian, 2007). The obvious difference between these two banking system is that they are both governed by different principles and banking act. Syariah compliance in Islamic banking tenets required the profit-loss sharing ratio is predetermined in advance, while the interest rate of conventional banking is determined by base-landing-rate which influenced by other macro-environmental factors such as inflation and recession (Batchelor, 2005).

As banks perform intermediation functions and generate deposit liabilities, subsequently, it influences the level of money and thus lifts the interest for depositors and investors to have informed information on bank performance. Following this, they are extensive studies that measure banks efficiency using DEA. The basic CCR DEA model was modified through linear programming equation to suit the objective and need of various researches.

3. CONCLUSION

1. This paper defined the effectiveness and efficiency of banking sectors. It further elaborates the efficiency which is interchangeably used with productivity. Four components of efficiency classification (i.e., technical efficiency, scale efficiency, price efficiency and allocative efficiency) is explained to describe the overall productivity. In addition, DEA as a tool to measure efficiency in banking sector is introduced.
with its basic models that are DEA CCR model with CRS assumption and DEA BCC model with VRS assumption. Also, two approaches in measuring bank efficiency are production and intermediation is being discussed in this paper.

Future banking efficiency study can employ DEA to get relative efficiency scores of sample banks. It can measures banks technical efficiency by assuming intermediation approach using basic DEA CCR model under CRS assumptions. The CRS assumption is used considering all banks or DMUs under observations are operating at optimal scale under perfect competition (Ahmad and Luo, 2010). This technique will reveal one set of efficiency score (CCR-efficient) for each DMU for each year under study. The method is adopted from Yahya et al. (2012), Hassan et al. (2009), Mokhtar et al. (2008) and Sufian (2007).

Furthermore, DEA BCC model with VRS assumption can be used to measure efficiency score of each DMU using the same input-output variables with the purpose to investigate the contribution of inefficiency. This process will give another set of efficiency score (BCC-efficient) for each DMU for each year under observations. The purpose of this technique is to measure source and level of inefficiency through decomposition of technical efficiency.

In addition, the CCR-efficiency scores obtained from the DEA can be used as dependent variables to test the effect of bank characteristics on efficiency score using linear regression. Prior to this, the sample banks (DMUs) under observation will be re-classified based on their bank characteristics. This method is adapted from Hassan et al. (2009).

**REFERENCES**


