Technical Efficiency of Commercial Banks in Malaysia: An Application of Window Data Envelopment Analysis

Wahidudin, Ahmad Nazri

2010

Technical Efficiency of Commercial Banks in Malaysia: An Application of Window Data Envelopment Analysis

Ahmad Nazri Wahidudin[§]

ABSTRACT

This paper analyzes technical efficiency of domestic commercial banks in Malaysia between 1995 and 2009 by using the Data Envelopment Analysis (DEA) window analysis. By this approach, the technical efficiency is analyzed sequentially with a certain window width (i.e. the number of years in a window) using a panel data of five domestic banks. The main idea is to capture the temporal impact on bank technical efficiency and see its short-run evolution from one window to another, in particular the pure technical efficiency (X-efficiency or managerial efficiency) and scale efficiency. By this, the study avoids the comparison of banks in different years as separate observations measured against each other, which can be unrealistic because of the significant technological diffusion in banking over the period under analysis.

The results suggest that on average the domestic commercial banks have some degree of inefficiency, which is more so due to pure technical rather than scale effects. Thus, Malaysian commercial banks should gain more from reducing the input quantities used or increasing the output quantities produced. The commercial banks should not worry too much about not choosing the correct scale for production though the study found out that on average the banks have not fully exhausted economies of scale.

Keywords: pure technical efficiency, scale efficiency, bank productivity, DEA.

1.0 INTRODUCTION

Dramatic changes have taken place in the Malaysian banking industry. From a number of 22 domestic conventional commercial banks as at the end of 1998, the number of banks reduced to ten only as at the beginning of the following year. As at the end of 2009 there are only nine domestic conventional commercial banks left. As local banking industry faces competitive incursion from beyond its shore, issues of bank efficiency becomes more important. Many studies were done to measure bank performance especially with regards to its efficiency, i.e. the degree to which the bank used its resources to produce outputs in comparison to the optimal used (best practice) of resources to produce the similar outputs. Given the changes in the Malaysian banking industry, a study on the commercial banks efficiency should be justifiable. We wish to see how efficiency evolves over time and whether there is any change in the banks efficiency levels.

In this paper, we consider the relative technical efficiency of five domestic commercial banks from 1995 to 2009. These banks existed in their enlarged merged form within that period. The study analyzes the bank efficiency using the Data Envelopment Analysis (DEA), which is a non-parametric, linear programming methodology. Charnes, Cooper and Rhodes (1978) developed the DEA model based on the earlier original work of Farell in 1975, in which they introduced a measure of technical efficiency focuses on input-reducing and assumed constant return to scale (CRS). Later Banker, Charnes and Cooper (1984) proposed a model that allows variable return to scale (VRS) to determine the technical efficiency devoid of scale effects, which decomposes the overall technical efficiency into pure technical efficiency and scale efficiency. The study also determines whether a bank suffers

[§] Affiliated with Business Programme Area, Faculty of Business & Computing. Institut Teknologi Brunei, Brunei Darussalam.

decreasing return to scale or increasing return to scale if scale inefficiency exists. The study analyzes the commercial bank efficiency using the DEA window analysis, by which the efficiency is analyzed sequentially with a defining window (i.e. a certain number of year periods) using a panel data of sample banks. The main idea is to capture the temporal impact on bank efficiency and see the short-run evolution of bank efficiency from one window to another. By this, the study avoids the comparison of banks in different years as separate observations measured against each other, which can be unrealistic because of the technological diffusion in banking over the period under analysis, would contribute to efficiency. The study approaches the analysis from two banking models, i.e. (i) a cost-revenue model in which a bank produces various services for its customers at certain costs, and (ii) an intermediation model in which a bank transfers funds from depositors to borrowers for profit.

In the next section, the paper discusses the efficiency concept and Data Envelopment Analysis models. Section 3 reviews the current literatures on banks using DEA. Section 4 describes the data and methodologies used in our empirical study. Section 5 presents and discusses the results of DEA window analysis and Section 6 concludes the paper.

2.0 DATA ENVELOPMENT ANALYSIS

Data Envelopment Analysis is a non-parametric method using a mathematical technique called linear programming (LP), which concerns with the allocation and utilization of limited resources. It is a mathematical process to optimize the value of certain output objective, for example maximize profit or minimize cost, when the input factors, for example labor, capital or raw materials, involved are subject to constraints. Charnes, Cooper and Rhodes (1978) first introduced DEA that measured efficiency based on Farrell's (1957) concept of efficiency measure. Charnes, Cooper and Rhodes (1978) developed a model which had an input orientation and assumed constant returns to scale. It is sometimes named as CCR model. Later, Banker, Charnes and Cooper (1984) proposed a model which assumes variable returns to scale. This model is also referred to as BCC model. The latter measures efficiency without the scale effect and the calculation of scale efficiency is possible by determining the difference between the two models' efficiency scores.

DEA is a performance measurement tool which evaluates the relative efficiency of production and service delivery entities in organizations. These entities are known as decision-making units (DMUs) in DEA. LP method solves a linear mathematical problem by generating a "virtual" efficient DMU and compares it with the observed DMU under analysis. The "virtual" efficient DMU is created from the DMUs which are found to efficient (best practice) and become benchmarks to the observed DMU. From this peer analysis, the degree of efficiency of the observed DMU is known and the slacks are identified and quantified. Slacks are the excess quantities of inputs (outputs) that can be reduced (increased) to achieve efficiency after all inputs (outputs) have been reduced (increased) in equal proportions to attain the best practice. DEA identifies the "best practice" DMUs and gives a perfect score of one (full efficiency) and any divergence from the "best practice" is considered inefficient. The degree of inefficiency depends on the score the DMUs received. The efficiency score based on multiple inputs and outputs is given by:

$$\text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}} \qquad (1)$$

A DMU that obtains a score of one is considered to be efficient and any score less than one indicates that it is inefficient.

2.1 EFFICIENCY CONCEPTS

The efficiency of an organization entails two main components: technical efficiency (TE) and allocative efficiency (AE). Technical efficiency reflects the ability of the organization to obtain maximum output from a given set of inputs. Allocative efficiency reflects the ability of an organization to use the inputs in optimal proportions, given their respective input prices. These two measures combined give a measure of total economic efficiency (TEE) or cost efficiency (CE). Technical efficiency can be decomposed into scale efficiency (SE) and pure technical efficiency (PTE).

Scale efficiency relates to the size of operation in the organization. The organization can take advantage by altering its operational size towards optimal scale and enjoys constant returns to scale. In other words, the relationship between output and inputs is constant in which the output changes equally in proportion to the changes in all inputs. The organization is said to be scale efficient. If output increases in proportion more than inputs, then the organization exhibits scale economies or increasing returns to scale (IRS). In practical term, the organization is under size and has ample room to take advantage from operational expansion and gains a higher proportional increase in output.

Conversely, if the proportional increase in output is less than the inputs, then the organization suffers scale diseconomies or decreasing returns to scale (DRS). In other words, the organization is overly large and has moved beyond the region of optimal scale. It does not benefit from the operational size growth as the increase in output is much less than the proportional increase in inputs. Hence in an organization, any divergences from the optimal scale is said to be scale inefficient.

The proportion of technical efficiency which is not attributed to the divergences from optimal scale is sometimes called as managerial efficiency or X-efficiency (pure technical efficiency). This relates to decision by a manager in formulating an input mix in order to generate output. There are many possible input mixes that attain various potential levels of output at the manager's disposal. Among the many, one input mix will be optimal and allows the organization exhibits pure technical efficiency. An organization can be both scale as well as pure technical inefficient. But only one is the major source of overall technical inefficiency.

DEA evaluates relative efficiency based on the premise that all efficient DMUs lay on a production frontier. The models consider N banks (DMUs), each using k different inputs and producing s different outputs. For ith DMU (i = 1,2,…,N) with an jth input (j = 1,2,…,K), x_{ij}, and an rth output (r = 1,2,…,S), y_{ir} represent the data of a DMU. The variables u_r and v_j are the output and input weights, respectively. Based on Charnes et al. (1978) the efficiency ratio (h) for DMU_0 is given by:

$$\max\ h_0(u,v) = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{j=1}^{k} v_j x_{j0}}$$

subject to

$$\frac{\sum_{r=1}^{s} u_r y_{ri}}{\sum_{j=1}^{k} v_j x_{ji}} \leq 1, i = 1,2,...,N$$

$$u_r, v_j \geq 0, \qquad (2)$$

where the first inequality ensures that the efficiency ratios for the other DMUs are less than one and the second inequality requires that the output and input weights are positive. The above ratio measure has infinite solutions and following Charnes et al. (1978) transformation of the denominator in the above efficiency ratio is set to equal to one which provides the transformed linear problem as:

$$\max\ h_0(u) = \sum_{r=1}^{s} u_r y_{r0}$$

subject to

$$\sum_{j=1}^{k} v_j x_{j0} = 1$$

$$\sum_{r=1}^{s} u_r y_{ri} - \sum_{j=1}^{k} v_j x_{ji} \leq 0, i = 1,2,..., N$$

$$u_r, v_j \geq 0. \qquad (3)$$

The dual linear programming of the problem is written as follows:

min (λ) θ_0
subject to

$$\sum_{i=1}^{N} \lambda_i y_{ri} \geq y_{r0}, r = 1,2,..., s$$

$$\theta_0 x_{k0} - \sum_{i=1}^{N} \lambda_i x_{ji} \geq 0, j = 1,2,..., k$$

$$\lambda_i \geq 0. \qquad (4)$$

The value of θ obtained is the efficiency score of the DMU_0. By solving the linear programming problem repeatedly, efficiency score of each DMU, i = 1, 2,…, N will be obtained and it will satisfy $\theta \leq 1$. The value of 1 indicates a point on the production frontier and hence a DMU is technically efficient. The above formulation assumes constant returns to scale (CRS) where the weights λ_i have no constraints other than a positive condition. Banker, Charnes and Cooper (1984) suggests an extension of the CRS specification to allow for variable returns to scale (VRS). It is necessary to add a convexity constraint where λ_i is equal to one and the input-oriented model can be written as follows:

min (λ) θ_0
subject to

$$\sum_{i=1}^{N} \lambda_i y_{ri} \geq y_{r0}, r = 1,2,..., s$$

$$\theta_0 x_{k0} - \sum_{i=1}^{N} \lambda_i x_{ji} \geq 0, j = 1,2,..., k$$

$$\sum_{i=1}^{N} \lambda_i = 1$$

$$\lambda_i \geq 0. \qquad (5)$$

BCC model provides technical efficiency scores greater than or equal to those obtained from the CCR model because it forms a convex hull that envelopes the data points more tightly than the latter's conical hull. Figure 1 illustrates this analysis in which a DMU at point T is fully efficient which it lays on the CRS frontier 0C, i.e. when TE scores of CRS and VRS are all equal to one. When VRS is considered, any DMU lays on the frontier RSTUV is pure technical efficient. That is when VRS TE score is larger than CRS TE score. But any DMU at the points S, U and V is not scale efficient.

As stated earlier, technical efficiency can be decomposed into scale efficiency (SE) and pure technical efficiency (PTE or non-scale efficiency). To determine the scale efficiency of a DMU, the difference between VRS TE scores and CRS TE scores is calculated. If VRS TE score is equal to CRS TE score, then the DMU is scale efficient and exhibits constant returns to scale. Otherwise, the DMU is scale inefficient and does not operating in the region of optimal scale or CRS. Thus, one problem with this measure is that it is not known whether the DMU is operating in the region of increasing returns to scale (IRS) or decreasing returns to scale (DRS). As a solution to this problem the DEA uses the nonincreasing returns to scale (NIRS) specification. The linear programming in (5) has to be solved with a constraint $\lambda_i \leq 1$. To determine whether a DMU is IRS or DRS, the VRS TE score with NIRS TE score are compared. If they are equal, then the DMU exhibits DRS and has gone beyond the region of optimal scale (CRS). If they are not equal, then the DMU exhibits IRS and has yet to reach the region of optimal scale in its operation. In summary, we have three efficiency measures each with its own specification, which can be expressed in ratios as follows:

CRS efficiency ratio = KL/KM
VRS efficiency ratio = KM/KN
SE ratio = KL/KM

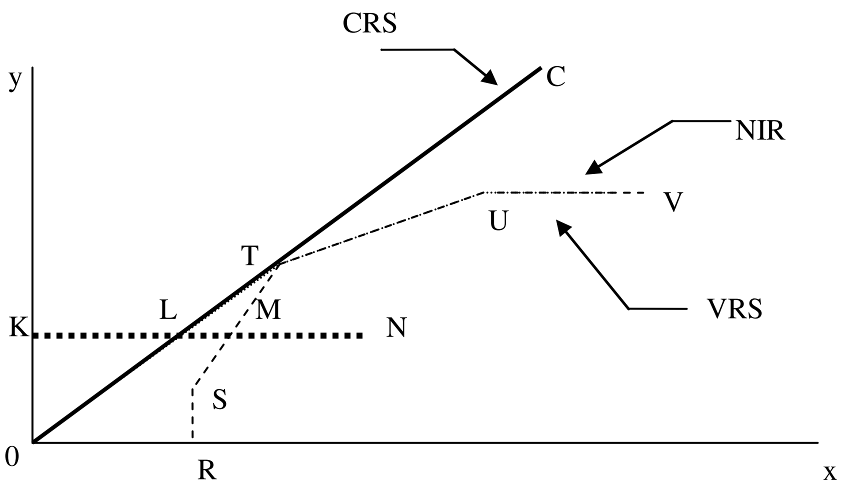

Figure 1: The Production Frontier and Returns to Scale

Referring back to Figure 1 the DMU at point L is technical efficient and the one at point M is pure technical efficient but scale inefficient. The DMU at point N is not technical efficient at all (both scale and non-scale). When CRS is assumed, the technical inefficiency covers the distance LN and when VRS is assumed, the pure technical inefficiency covers the distance MN. Its scale inefficiency is the distance LM. The DMU at point S has scale inefficiency and its exhibits increasing returns to scale with its low output level. The DMUs at points U and V have decreasing returns to scale, with the DMU at point V being the furthest from optimal scale.

3.0 BRIEF LITERATURE REVIEW

A number of studies have applied DEA to examine the efficiency of banking institutions. A few have used DEA window analysis approach to evaluate banks, which uses time-series data. Yue (1992) evaluates 60 largest Missouri banks using the window approach for the period 1984 to 1990. He examines the banks from intermediation model perspective using four inputs (i.e. interest expense, non-interest expense, savings deposits, and demand deposits) and three outputs (i.e. interest income, non-interest income and total loans). He reports that the major source of inefficiency comes from pure technical inefficiency and larger banks are scale efficient. Miller and Noulas (1996) evaluate 201 large US banks from 1984 to 1990. Employing four inputs (i.e. interest expense, non-interest expense, savings deposits, and demand deposits) and three outputs (i.e. various types of loans, interest income and non-interest income), they approach the analysis from the intermediation model. They find that larger and profitable banks are more scale efficient and those that are inefficient tend to operate under decreasing returns to scale. The bulk of inefficiency however traces to the pure technical inefficiency.

Paradi et al (2001) analyzes five largest Canadian banks over a twenty-year period from 1981 to 2000. Using DEA window analysis they examine the banks from two banking models, i.e. the production model and intermediation model. In the production, they employ four inputs which entail interest expense, non-interest expense, fixed assets and number of employees. They employ five outputs which entail non-interest income, total deposits, other banks' total deposits, total loans and marketable securities. In the intermediation model, they employ seven inputs which entail non-interest income, total deposits, debenture/subordinated debts, other liabilities, shareholders funds, fixed assets and number of employees, and five outputs which entail non-interest income, total loans, marketable securities, deposits with BOC and non-interest earning assets. The study finds that both models give similar DEA results and report that bank technical efficiency changes over time and from one year to another.

Jemric and Vujcic (2002) evaluate Croatian banks in the period from 1995 to 2000. They analyze technical efficiency using both production and intermediation models. In the former they employ four inputs (i.e. interest expense, non-interest expense, personnel costs and capital expenditures) and two outputs (i.e. interest income and non-interest income). In the latter they employ three inputs (i.e. total deposits, fixed assets and number of employees) and two outputs (i.e. total loans and money-market securities). They find that the most technical efficient banks are either the smallest or the largest banks. The smallest banks are often niche banks, and have no excess labor and have low costs of fixed assets. The major source of inefficiency comes from the pure technical efficiency.

Drake and Hall (2003) examine bank efficiency in 149 different types of Japanese banks for the financial year ending March 1997. The data covers all sizes of banks. From intermediation approach, the study uses three inputs (i.e. non-interest income excluding personnel costs, total deposits and fixed assets) and three outputs (i.e. non-interest income, total loans and marketable securities). The study finds that on average big banks are technical efficient but ones that are inefficient tend to exhibit decreasing returns to scale. The major source of inefficiency comes from the pure technical inefficiency and small banks are more so exhibited such inefficiency. Webb (2003) examines UK's large retail banks based on a DEA window analysis for the period from 1982 to 1995. He finds that the banks average efficiency deteriorates over tine and majority of the banks exhibit scale inefficiency. The bigger ones tend to operate under decreasing returns to scale.

Sufian (2004) examines Malaysian commercial banks during the merger year, pre- and post-merger period. The data covers a period between 1998 and 2003 of ten commercial banks. The banks show a high overall efficiency of about 96%, and small and medium size banks enjoy greater scale efficiency. Large banks do not benefit from any scale economies from the mergers. The results indicate that these large banks should shrink in size to benefit from their economies of scale. On the whole the study finds out that scale inefficiency dominates pure technical efficiency in Malaysian commercial banks.

Abdul Majid and Sufian (2005) analyze Malaysian commercial banks in the post-merger, i.e. after 2001. Their results suggest that large banks do not benefit from any scale effects and only smaller size banks enjoy such benefit. On the whole the banks' inefficiency is attributed to the pure technical efficiency as their second-stage analysis indicates that risk input has a greater influence on X-efficiency.

Chambers and Cifter (2006) investigate the productivity of Turkish banks which covers a period from 2002 to 2004. The study uses five input variables: branch numbers, number of branch staff, total assets, total loans, and total deposits. The output variables are net profit/loss, return on equity, net interest income, and non-interest income. The study finds out that the banks are technically inefficient which does not come from scale economies. The banks do not enjoy any benefit from scale economies.

4.0 DATA AND METHODOLOGY

The sample used in the study consists of data from five commercial banks that exist until today albeit in their enlarged form. The sampling based on two criteria, i.e. the listing of banking institutions legally merged and listed as at December 31, 2001 published in BNM Monthly Statistical Bulletin, and the availability of bank financial statements from 1995 to 2009 in the public domain. Based on these two criteria we acquired a sample of five commercial banks including one having the smallest asset size of RM28.5 billions and one having the largest asset size of RM238.9 billion as at the end of their financial year in 2009. The banks' balance sheets and income statements provide all the required data for the analysis. We approach the study from two commonly used banking models, i.e. a production model and intermediation model.

A production model views banks as entities providing products and services to customers using various resources. The products and services, e.g. the various types of loans and deposits, are outputs and the resources, e.g. labor, capital and overheads, are inputs in this model. This approach measures cost efficiency because it considers the operating costs of banking. We use a cost-revenue variant model that entails interest income and non-interest income as outputs, and interest expense, personnel cost and non-personnel cost (other overheads) as inputs.

An intermediation model views banks as financial intermediaries who mobilize funds from the surplus units and lend them out to the deficit units in the economy. These mobilized funds, i.e. savings and deposits, and the costs incurred and the assets used in the intermediation process are inputs to the model. The ways in which the funds can be lent out or invested are considered outputs. This approach measures the organizational efficiency and the economic viability of banks because it considers all the costs of banking. We use total deposits, non-interest expense (overheads) and fixed assets as the inputs, and total loans, marketable securities and non-interest income as the outputs to the model. We wish to include the number of employees of each bank in the sample as inputs but they are not readily available in the public domain. In lieu of this, we use the bank overhead expenses as one of the inputs. The fixed assets represent the amount of plant, property and equipment as stated in bank balance sheets. The marketable securities are those securities invested and traded as stated in bank balance sheets. The underlying idea is to capture the measures for the banks' labor, capital, operating costs and revenues.

We use DEA window analysis based on an input orientation model to measure bank efficiency in this study. The study should be able to see the bank efficiency evolves over time and to see whether any size effect exists in the bank efficiency. The results will be discussed from time as well as individual bank perspective. The data spans for fifteen years and forms thirteen windows, and each window covers a period of three years. The choice for the window width agrees with the original work in Charnes, Clark, Cooper and Golany (1985). Each window is analyzed independently but the years in each window move into the next window based on the principle of moving averages. Thus, this study generates several short-run analysis of technical efficiency. Otherwise, it would be a single longitudinal analysis of a panel data of fifteen years from five banks. We believed technology has played a very significant role in the efficiency of banks over time. By using window analysis, the study should be able to minimize the technological impact on the technical efficiency of banks. Using the window analysis approach is useful in increasing the number of data points because we have a small sample size. Nunamaker (1985) suggest that the sample size should be triple the sum of outputs and inputs. At least the sample size should be greater than the product of outputs and inputs (Talluri 2000). Thus, we have increased our data points since we have thirteen windows each with a window width of three years and consist of five commercial banks. A descriptive statistics of the variables used in the analysis are given in Table 1, where (O) or (I) indicates whether a variable enters as an input and output in each respective model.

Table 1: Descriptive Statistics of the Input and Output Variables

Production Model	Interest Expense (I)	Staff Cost (I)	Non-Staff Cost (I)	Interest Income (O)	Non-Interest Income (O)
Mean	1363.6	374.1	367.9	3776.2	887.9
S. Error	159.0	47.1	45.6	1346.7	283.5
Median	789.0	204.4	206.4	1337.0	231.0
S. Deviation	485.8	209.1	#N/A	936.6	4602.1

Minimum	1377.3	408.1	395.3	11662.8	2455.3	
Maximum	75.3	21.0	37.0	116.9	11.6	
Intermediation Model	Deposits (I)	Non-interest Expense (I)	Fixed Assets (I)	Total Loans & Leases (O)	Marketable Securities (O)	Non-Interest Income (O)
Mean	43544.9	742.4	492.6	32439.8	7045.8	646.8
S. Error	5622.6	91.0	74.1	4272.1	1050.1	109.4
Median	23979.6	408.9	287.6	13964.3	3550.7	231.0
S. Deviation	48693.5	787.7	641.9	36997.8	9094.3	947.6
Minimum	1748.6	58.0	25.6	1357.0	239.0	11.6
Maximum	193574.8	3736.6	3891.2	144431.8	47020.2	4602.1

5.0 RESULTS AND ANALYSIS

We find that the analysis of both cost-revenue variant and intermediation models give similar account. This agrees with the evidence produced by Paradi et al. (2001) in their studies of Canadian banks using both types of banking models. Here we present the analysis of intermediation model. Table 2 presents the mean efficiency scores of each window from Window 1 to Window 13. The CRS and VRS mean efficiency scores in windows 1 through 9 are less than one, but nonetheless the scores are considerably high. Since the scores are less than one, the banks have considerable technical inefficiency. In every window, the bulk of the inefficiency is due to the banks' failure to minimize the input amounts used in the delivery of services. In other words, the banks have pure technical inefficiency which indicates that the banks consumed resources more than optimally required (best practice) to produce the given outputs.

Similar results found by Yue (1992), Jemric and Vujcic (2002), and Drake and Hall (2003) Webb (2003), Abdul Majid and Sufian (2005) and Chambers and Cifter (2006). For example in window 9 (the period between 2003 and 2005), the VRS mean efficiency score is 0.9745 which also implies that on average the banks have 2.55 % inefficiency. In order to operate at the efficiency frontier the banks have to reduce all input amounts by the same proportion without affecting the current output levels. After reducing all the inputs proportionally and to avoid further wastage and inefficiency relative to the best practice, the banks can reduce further input amounts (i.e. the input slacks). During the period defined as window 9, the major input slack is in the fixed assets (plant, property and equipment). Another input slack is the non-interest expense (overheads). This could be attributed to over the years increased in the personnel expenses as staff salary and benefits account for at least 50% of the overhead expenses in the banking industry. The variation in the mean scores is larger from window 5 to window 8. This could be attributed to the post-consolidation effect as the banks trying to deal with newly enlarged resources and higher output levels. We can see that the variation is lower from window 9 onwards as the impact has leveling off.

Table 2: CRS, VRS, and SE Mean Efficiency Scores of Windows

Windows	CRS		VRS		SE
	Mean	Std Dev	Mean	Std Dev	Mean
1	0.933	0.072	0.990	0.023	0.943
2	0.964	0.050	0.987	0.026	0.977
3	0.970	0.044	0.982	0.036	0.987
4	0.949	0.063	0.974	0.041	0.975
5	0.935	0.108	0.962	0.082	0.972
6	0.948	0.109	0.963	0.087	0.984

7	0.952	0.099	0.963	0.089	0.989
8	0.933	0.101	0.967	0.088	0.965
9	0.956	0.073	0.975	0.056	0.981
10	0.975	0.052	0.053	0.047	0.992
11	0.959	0.063	0.065	0.050	0.980
12	0.957	0.053	0.056	0.042	0.980
13	0.968	0.048	0.050	0.008	0.971

From Table 3, we also find that on the average, the banks have scale economies (IRS) and have not grown too big in their operation. However, the primary inefficiency is traced to the pure technical effects rather than an incorrect production scale. Thus, the banks will not gain much more from any action in finding the perceived correct production scale. From DMU (i.e. individual bank) perspective, on average DMU 4 is scale inefficient. It has exhausted its scale economies and gone beyond the region of optimal scale (DRS). We find that bigger banks tend to exhibit DRS or have gone too far in operational scale and grown overly large. We test for significance by which the Spearman's coefficient of rank correlation of 0.392 is significant at 1% level. This evidence agrees with the findings of Miller and Noulas (1996), Drake and Hall (2003), Webb (2003), and Sufian (2004). DMU 2 and 3 have technical inefficiencies that are traced to scale effects too. But they have not exhausted their scale economies (IRS), hence they will benefit more from size growth. DMU 1and 5 have technical inefficiencies that are primarily due to pure technical reasons. They will benefit more from reducing input amounts so as to move to the efficient frontier. However, we wish to highlight DMU 5 which has the highest variation in its mean efficiency scores. We find that after its consolidation exercise in the period 2000 - 2001, the input trend for overhead expenses increased by 46.5 % from the previous financial year. The personnel cost, which increased by 63% from the previous financial year, contributed significantly to the sudden spike in the input trend. DEA measure is very sensitive to such data trend. We can see that DMU 5 has the lowest mean efficiency scores.

Table 3: CRS, VRS, and SE Mean Efficiency Scores of DMUs

	CRS		VRS		SE
DMU	Mean	Std Dev	Mean	Std Dev	Scale Eff.
1	0.946	0.057	0.969	0.049	0.976
2	0.967	0.043	0.989	0.027	0.979
3	0.976	0.049	0.994	0.020	0.982
4	0.973	0.046	0.998	0.011	0.975
5	0.907	0.123	0.925	0.108	0.980

We test for bank size effect on the mean efficiency scores and find that it has no correlation with the technical, scale or pure technical efficiency. Under the cost-revenue model, the bank asset size does correlate with scale efficiency. Drake and Hall (2003) found evidence in Japanese banks that big banks have high level of scale efficiency. Our results show significance at 5% level though the Spearman's coefficient of 0.250 is considered very low. We also find that the bank efficiency fluctuates over time because a bank that has scale efficiency in one window may have pure technical efficiency in another window. Yue (1992) and Paradi et al. (2003) found similar evidence in Missouri banks and Canadian banks, respectively. We observe that the variation in mean efficiency scores get higher in the period after 1999 and later get lower in the period after 2003. The variation is greater in the intermediation model than the cost-revenue model. Jemric and Vujcic (2003) found similar evidence in the efficiency analysis of Croatian banks. The cost-revenue model uses input and output variables from the bank income statement alone. The intermediation model uses variables both from the balance sheet and income statement, which captures two facets of the bank operation. The banks then were dealing with newly enlarged resources and higher given output levels. Hence, we see the intermediation model exhibits greater variability in its mean efficiency scores.

6.0 CONCLUSIONS

This study has examined five domestic commercial banks in Malaysia over the period from 1995 to 2009. By using DEA window analysis we inflate the number of observed data points given our very small sample size. In each window, the number of banks is tripled because each bank at a different year is taken as independent DMU. Thus, we obtained information about the short-run evolutions of DEA efficiencies of every bank during the fifteen years. Several conclusions emerge from this study. Firstly, the commercial bank mean efficiency scores or rankings have no correlation with the bank asset size. Small commercial banks may be equally inefficient as large commercial banks. However, there is evidence that big banks tend to be scale efficient under the cost-revenue model but the correlation is very low though statistically significant. Secondly, big banks tend to exhaust their scale economies and experience decreasing returns to scale. Small banks have not exhausted their scale economies and may benefit from size growth. Thirdly, on average the commercial banks have more pure technical inefficiency rather than scale inefficiency. Thus, the banks should gain more from reducing input quantities to produce the given outputs and to operate at the efficient frontier. Lastly, the technical efficiency of commercial banks fluctuates over time. A bank can be scale efficient in one period and may not be efficient in another period.

There are other factors that influence bank efficiency and some of these factors are beyond the control of bank managers. For instance, a bank may have a certain socio-economic objective to fulfill and this may influence its potential efficiency score. The extents to which a bank taking risk may translate into higher efficiency scores as the bank produce more loans. But the loans may not be necessarily of good quality and may have a higher probability of default. Second-stage econometric or regression analysis can be done to test for some of the factors for their explanatory powers.

REFERENCES

1) Abdul Majid, Mohamed Zulkhibri and Sufian, Fadzlan, "Post-merger Banks' Efficiency and Risk in Emerging Market: Evidence from Malaysia", *The IUP Journal of Bank Management*, Vol. IV, No. 4, November 2005, pp. 16 - 37

2) Banker, R.D., Charnes, A., Cooper, W.W., "Models for Estimating Technical and Scale Efficiencies", *Management Science*, Vol. 30, 1984, pp. 1078-1092.

3) Chambers, Nurgul and Cifter, Atilla, "The Effect of Scale on Productivity of Turkish Banks in the Post-Crises Period: An Application of Data Envelopment Analysis", Traditional International Finance Symposium, Marmara University, 2005.

4) Charnes, A., Cooper, W.W., Rhodes, E., "Measuring Efficiency of Decision Making Units", *European Journal of Operational Research*, Vol. 1, 1978, pp. 429-444.

5) Drake, L., Hall, M.J.B., "Efficiency in Japanese Banking: an Empirical Analysis", *Journal of Banking & Finance*, 27, 2003, pp. 891-917.

6) Jemric, I., Vujcic, B., "Efficiency of Banks in Croatia: A DEA Approach", *Comparative Economics Studies*, XLIV, No. 2, 2002, pp. 169-193.

7) Kamaruddin, Badrul Hisham, Safa, Mohd Samaun and Mohammad, Rohani, "Assessing Production Efficiency of Islamic Banks and Conventional Banks Islamic Window in Malaysia", *International Journal of Business and Management Research*, 1(1), 2008, pp. 31 – 48.

8) Miller, S.M., Noulas, A.G., "The Technical Efficiency of Large Bank Production", *Journal of Banking & Finance*, 20, 1996, pp. 495-509.

9) Paradi, J.C., Asmild M., Aggarwall, V., Schaffnit, C., "Performance Evaluation in an Oligopoly Environment: Combining DEA Window Analysis with the Malmquist Index Approach – A Study of the Canadian Banking Industry", Centre for Management of Technology and Entrepreneurship, University of Toronto, 2001.

10) Peristiani, S., "Do Mergers Improve the X-Efficiency and Scale Efficiency of U.S. Banks? Evidence from the 1980s", *Journal of Money, Credit and Banking*, Vol. 29. No. 3, 1997, pp.326-337.

11) Sufian, Fadzlan, "The Efficiency Effects of Bank Mergers and Acquisitions in a Developing Economy: Evidence from Malaysia", *International Journal of Applied Econometrics and Quantitative Studies*, Vol. 1, No. 4, 2004, pp. 53 – 74.

12) Talluri, S., "Data Envelopment Analysis: Models and Extensions", Decision Line, May 2000, pp. 8-11.

13) Webb, R.M., "Levels of Efficiency in UK Retail Banks: a DEA Window Analysis", *International Journal of the Economics of Business*, Vol. 10, No. 3, 2003, pp. 305-322.

14) Yue, P., "Data Envelopment Analysis and Commercial Bank Performance: A Primer with Applications to Missouri Banks", Review – Federal Reserve of St. Louis, Vol. 74, No.1, 1992, pp.31-47.