Measuring Efficiency of Foreign Banks in the United States

Joon J. Park
Associate Professor, Department of Business Administration
University of Arkansas at Pine Bluff
1200 North University Drive, Pine Bluff, Arkansas 71601, USA
E-mail: parkj@uapb.edu
870) 575-8083

Abstract
The number of foreign banks having branch offices in many leading metropolitan areas began to grow. To date it has largely found that foreign banks are less efficient than domestic banks. This paper investigates the factors that determine differences in efficiency of foreign banks in US market. The impact of home, host, and parent bank factors are considered within the frameworks offered by DEA.

Keywords: Banks, Data envelopment analysis, productivity, efficiency, Malmquist
1. Description of the primary issues
The increase in international trade flows, foreign investment activities, international migrations and the globalization has brought a substantial inflow of foreign banks to the United States. The number of foreign banks having branch offices in many leading metropolitan areas began to grow. To date it has largely found that foreign banks lease efficient than domestic banks (Berger et al, 2000). This paper investigates the factors that determine differences in efficiency of foreign banks in US market. The impact of home, host, and parent bank factors are considered within the frameworks offered by DEA.

2. Review of the literature

Berg et al. (1992) showed that the mean productivity index was 1.09 between Finland and Norway, 1.52 between Finland and Sweden, and 1.40 between Norway and Sweden. Sathye (2002) studied the productivity changes in the Australian banking system over the period 1995–1999 and found that the mean total factor productivity was 1.013. Kirkwood and Naum (2006), a study on the cost efficiency of Australian in producing baking services and profit, found that the major banks have improved their efficiency in producing banking services and profit, while the regional banks have experienced little change in the efficiency of producing banking services, and a decline in the efficiency of producing profit.

3. Data
3.1. Data
The data are obtained from financial reports of U.S. Securities and Exchange Commission (SEC) and Federal Deposit Insurance Corporation (FDIC) and adjusted in 2007 dollars. The home factor data of foreign banks are obtained from corresponding website(s) of the central banks of home countries.

3.2. Variables: Selection of Inputs and Outputs
Measures of efficiency vary according to the selection of inputs and outputs. Therefore, the selection process of inputs and outputs in empirical analysis is an important step. In general, the desirable number of inputs and outputs is less than one-third of the overall DMUs to increase the discriminating power of efficiency. However, there are no clear
criteria on calculating the efficiency values by selecting the suitable outputs and inputs. Using correlation analysis among variables within each input and output to eliminate the factors having the highest correlation classifying the DMUs according to their main roles are done frequently.

Banks could be classified according to their roles, for example, including the functions of providing productivity, intermediation and added value. When researchers examine input and output products while considering each function in detail, the focus is on the ability to provide savings and loan services using labor and capital. Thus, inputs could be labor, capital and other expenses, whereas outputs could be deposits and loans. If loaning capital from net lenders to net borrowers is the primary role of banks in the intermediation function approach, for instance, the major inputs considered include deposits, and the output is the amount of funds loaned. On the other hand, the value-added approach views banks as the providers of financial services, and inputs and outputs are determined according to the pure contribution to profits from financial products.

Although these approaches could provide the basic theory of selecting inputs and outputs, they could never be absolute standards. The selection of input and outputs is achieved in various ways according to the purposes of the previous studies. The model estimated for this paper adopts a variation of the intermediation approach for measuring the efficiency of financial intermediaries and views a bank as using noninterest expenses (NIE), deposits (D), stock holders equities (SHE), and total debts (TD) to produce incomes before income tax (IBIT), investment securities (IS), and total loans (TL).

4. Methodology: DEA and Malmquist Index
DEA was introduced by Charnes, Cooper, and Rhodes (or CCR) (1978) and later extended by Banker, Charnes, and Cooper (or BCC) (1984). The CCR model has an input orientation and assumes constant returns to scale (CRS). The BCC model is an extension of the CRS DEA model and accounts for variable returns to scale (VRS) situations. Use of these two models allows the decomposition of technical efficiency into pure technical efficiency and scale efficiency.

Malmquist (1953) used a quantity index for use in consumption analysis. Caves et al. (1982) adapted Malmquist’s idea for production analysis and they named their productivity change indices after Malmquist. Färe et al. (1989) showed that the Malmquist indices can be decomposed into two components: technical efficiency change and technical change. The next section summaries CCR and BCC models; the Malmquist index, used frequently for vertical analysis, is discussed subsequently.

4.1. CCR Models
The CCR model uses linear programming to measure the relative technical efficiency of each decision-making unit (DMU) incorporating multiple inputs and outputs. The efficiency scores of DMUs fall between 0 and 1; fully efficient banks will have an efficiency score of 1. Suppose we have a set of $n$ peer DMUs that produce observed multiple output vector $Y$, by utilizing observed multiple input vector $X$, respectively. Then, the production possibility set $F$ is defined as Equation (1).

$$F = \{ (Y, X) | X \text{ can produce } Y \}$$

An efficient frontier (or production technology) can be represented with a set of DMUs that satisfy Pareto efficiency conditions among the production possibility set. This
efficient frontier requires following two basic assumptions (Shephard, 1970). First, the
efficient frontier should be satisfied with the convexity assumption of production
possibility set \( F \). The convexity assumption means that, for a DMU with a single input \( A \)
and single output \( B \), respectively, if \( (y^A, x^A) \in F \) and \( (y^B, x^B) \in F \), then
\[ (\lambda y^A + (1 - \lambda) y^B, \lambda x^A + (1 - \lambda) x^B, 0 \leq \lambda \leq 1) \in F. \] Second, the efficient frontier should be
satisfied with a free disposability assumption of inputs and outputs. The free disposability
assumption means that, for inputs, if \( (y^A, x^A) \in F \) and \( x^B \geq x^A \), then \( (y^A, x^B) \in F \),
and, for outputs, if \( (y^A, x^A) \in F \) and \( y^B \leq y^A \), then \( (y^B, x^A) \in F \).

Shephard (1970) provided another functional representation of production
technology as a definition of distance function:

\[
D(Y, X) = \min \left\{ \theta(Y, X / \theta) \in F \right\} \tag{2}
\]

where \( D(Y, X) \) is the output oriented distance function. This set can be described
mathematically by its sections. Therefore, the input oriented distance function is defined as
\[ \max \left\{ \theta(Y, X / \theta) \in F \right\}. \] To estimate of such a distance function, Aigner and Chu
(1968) introduced a nonparametric technique called linear programming. Later Charnes,
Cooper, and Rhodes (1978) developed a DEA methodology in which optimal solutions is
the reciprocal of Farrell’s (1957) technical efficiency estimates, given as:

\[
\text{Min } \theta - \varepsilon \sum_{r=1}^{n} s^+_r - \varepsilon \sum_{i=1}^{s} s^-_i
\]

s.t. \[
x_{ij} \theta - \sum_{j=1}^{n} x_{ij} \lambda_j - s^-_i = 0, \quad i = 1, 2, \ldots, m \tag{3}
\]

\[\sum_{j=1}^{n} y_{ij} \lambda_j - y_{ij} - s^+_j = 0, \quad r = 1, 2, \ldots, s \]

\[\lambda_j, s^-_i, s^+_j \geq 0, \quad \forall j, r, i\]

where we assume \( n \) units, each using \( m \) inputs to produce \( s \) outputs. We denote by \( y_{ij} \)
the level of the \( r \)th output \((r = 1, 2, \ldots, s)\) from unit \( j \) \((j = 1, 2, \ldots, n)\) and by \( x_{ij} \)
the level of the \( j \)th input \((j = 1, 2, \ldots, m)\) to the \( j \)th DMU.

Notable, \( \varepsilon \) is a very small positive number that prevents the weights from
vanishing (formally, \( \varepsilon \) should be seen as a non-Archimedean constant), \( s^-_i \) and \( s^+_j \)
represent the slack variables, and \( \lambda_j \) are variables whose optimal values will define an
efficient production possibility by minimizing inputs DMU_0 without detriment to its
output levels. As a result, the optimal solution of \( \theta \) represents the estimated efficiency
of DMU_0.

Equation (3) is called as CCR model, which adds a constant returns to scale
(CRS) condition of the efficient frontier to the two basic assumptions mentioned earlier.
The CRS condition means that, for \( k > 0 \), if \( (Y, X) \in F \), then \( (kY, kX) \in F \).

4.2. BCC Model and Scale Efficiency
Another DEA model, which is usually referred to as the BCC model is proposed by
Banker, Charnes, and Cooper (1984). This model is expressed by adding convexity
constraint such as $\sum_{j=1}^{n} \lambda_j = 1$ to the traditional CCR model. As a result, BCC model can separately estimate pure technical efficiency and scale efficiency on the assumption that variable returns to scale in production technology exist. If there is the BBC technical efficiency score is higher than the CCR technical efficiency score for a particular DMU, then this indicates that the DMU has scale efficiency, and that the scale efficiency is equal to the ratio of the CCR technical efficiency score to the BBC technical efficiency score.

4.3. Malmquist Productivity Index

The purpose of a Malmquist-type productivity index is to measure changes in productivity and efficiency over time. The basic idea behind the Malmquist index is to exploit the relation between distance functions of different time periods as ratios. The measured values that the Malmquist indices provided include efficiency changes (pure efficiency changes and technology changes), scale efficiency changes, and changes of total factor productivity.

Caves, Christensen, and Diewert (1982) defined an Malmquist productivity index as a ratio of distance function between periods $t$ and $t+1$ as depicted in Equation (4). This index is based on output oriented Malmquist index.

$$M^{t,t+1} = \frac{D(Y^t_j, X^t_j)}{D(Y^{t+1}_j, X^{t+1}_j)}, \quad j = 1, 2, \ldots, n$$

(4)

Fare, Grosskopf, Lindgren and Roos (1995) showed this index can be solved using linear programming. They then employed the geometric mean of the two output oriented Malmquist indices that expand above a measure of productivity change, as:

$$M^{t,t+1} = \left[ \frac{D'(Y^t_j, X^t_j)}{D'(Y^1_j, X^1_j)} \cdot \frac{D^{t+1}(Y^t_j, X^t_j)}{D^{t+1}(Y^1_j, X^1_j)} \right]^{1/2}$$

(5)

$$= \frac{D^{t+1}(Y^t_j, X^t_j)}{D^t(Y^1_j, X^1_j)} \cdot \left[ \frac{D'(Y^t_j, X^t_j)}{D'(Y^1_j, X^1_j)} \cdot \frac{D^t(Y^t_j, X^t_j)}{D^t(Y^t_j, X^t_j)} \right]^{1/2}$$

$t = 1, 2, \ldots, T-1, \quad j = 1, 2, \ldots, n$

where $D$ represents the distance function and $M$ is the Malmquist productivity index. Each DMU is identified by its input-output bundle $Y_j, X_j$ with a superscript indicating whether it is observed at time $t$ or $t+1$. The distance is defined as $D'$ or $D^{t+1}$ depending on whether the reference frontier is that of time $t$ or $t+1$. A value of $M$ greater than 1 (i.e. $M > 1$) denotes productivity growth, while a value less than 1 ($M < 1$) indicates productivity decline; $M=1$ indicates no productivity change.

The first term on the right-hand side in Equation (5) is a ratio of two distance functions; it measures the magnitude of the change in technical efficiency between $t$ and $t+1$. It is greater than, equal to or less than 1 if the DMU is moving closer to, unchanging, or diverging from the production frontier. The second term measures the technical change in the efficient frontier between $t$ and $t+1$. It measures the technological improvements between periods and is greater than, equal to or less than 1 when technological best practices are improving, unchanged, or deteriorating, respectively. The overall index of productivity is given by the product of the two components, so that an index bigger than 1 indicates that total factor productivity has increased, and vice versa for a value smaller
If variable returns to scale (VRS) are assumed, the first term on the right-hand side in Equation (5) can be further decomposed into two parts: one is pure technical efficiency change, which isolates the technical efficiency catching-up of units against the VRS technology frontier; the other is scale efficiency change that captures the movements towards or away from the constant return facet of the frontier. There are many different ways to measure the distance function that makes up the Malmquist productivity index. In the empirical part of this study (see section 4), we use the data envelopment analysis program (DEAP) (Coelli, 1996) to construct Malmquist indices.