Chapter 4

Theoretical Framework and Methodology

This chapter describes the method of nonparametric techniques of efficiency measurement Data Envelopment Analysis (DEA). Technical efficiency refers to the ability to minimize input use in the production of a given output. Allocative efficiency refers to the ability to use inputs in the appropriate nation of a given price of input. The efficiency measurement can be divided into input-orientated productive efficiency measurement and output-orientated productive efficiency measurement.

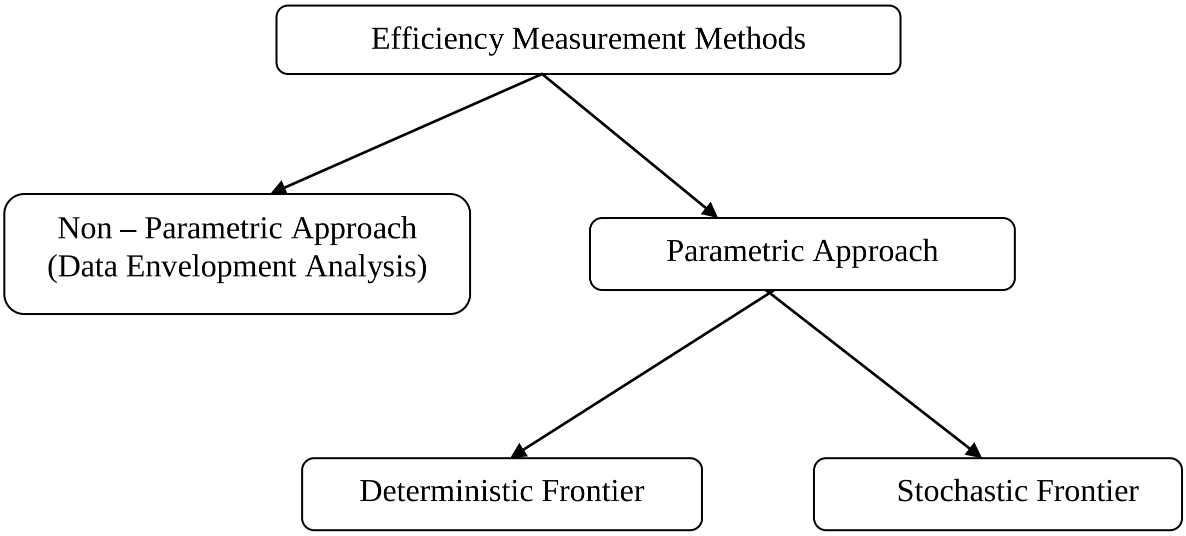
As shown in Figure 4.1, the efficiency measurement methods can be divided into two approaches which are non-parametric approach and parametric one.

The main method of non-parametric is Data Envelopment Analysis (DEA). DEA is a non-parametric method that uses piecewise linear programming to calculate efficiencies or best-practice frontiers in a given set of decision-making units (DMUs) such as firms. DEA does not require a hypothesized functional form linking input and outputs.

Parametric approach consists of deterministic frontier and stochastic frontier. These methods assume that a set of organizational units is defined to be assessed and that a single input is used to produce the outputs. However, a main problem of this approach is that it cannot cope with multiple inputs and multiple outputs. These problems can be overcome by the non-parametric method of comparative performance measurement.

Figure 4.1

Efficiency Measurement Method



4.1 The Measurement of Efficiency

The efficiency measurements are divided into input-oriented measurement and output-oriented measurement. This section will explain the input-oriented measurement. Output-oriented measures as the complement to the input-oriented measure will be discussed subsequently.

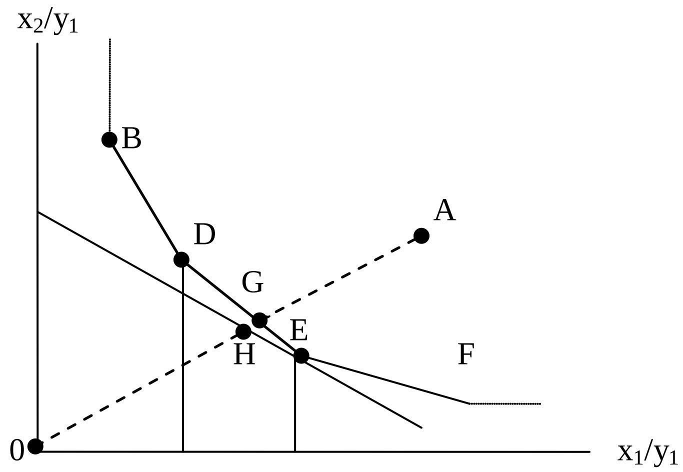
4.1.1 Input-Oriented Measurement

The efficiency measures will be defined with reference to input and output levels disregard to their prices. Efficiency defined in this manner is known as “technical” efficiency. Because technical efficiency does not reflect relative input prices or output values, it cannot give the best account of the performance.

Farrell illustrated the ideas of using a simple example involving firms, using two inputs (x_1 and x_2) to produce a single output (y), under the assumption of constant return to scale.

Figure 4.2

Contrasting Technical and Input Allocative Efficiencies



Source: Ray. (2004).

When input prices are known then the input allocative efficiency of a DMU can be measured, introduced by Farrell (1957) as price efficiency. Input allocative efficiency reflects the distance of the input mix used by a DMU from the optimum mix. It can be used to minimize the cost of output, in light of input prices. Input allocative efficiency complements the measure of technical efficiency that introduced initially.

The measure of input allocative efficiency can be illustrated with reference to DMU A in Figure 4.2. Where DMU A becomes technically efficient under its chosen input mix, it would operate at G. Thus,

$$\frac{0G}{0A} \text{ is the input technical efficiency of DMU A.}$$

The fraction to which the aggregate cost of the inputs at G could be written as $\frac{0H}{0G}$ and can be defined as,

$\frac{0H}{0G}$ is the input allocative efficiency of DMU A.

The fraction to which the aggregate cost of the inputs at A could be written as $\frac{0H}{0A}$ and can be defined as,

$\frac{0H}{0A}$ is the input overall efficiency of DMU A.

It could be written as,

$$\frac{0H}{0A} = \frac{0G}{0A}\frac{0H}{0G}$$

Therefore:

Input Overall Efficiency = Input Technical Efficiency x Input Allocative Efficiency

4.1.2 Output-Oriented Measurement

Output-oriented measurement as opposed to the input-oriented measurement will be presented in this section. Farrell illustrated the ideas of using a simple example involving firms, using two outputs (y_1 and y_2) to produce a single input (x), if the input quantity fixed at a particular level is held. Figure 4.3 shows the technology by a production possibility curve in two dimensions where the line ZZ′ is the production possibility curve. An inefficient firm is operating at point A lies below the curve, because ZZ represents the upper bound of production possibilities. The distance 0A/0B represents output-oriented technical inefficiency. That is, the amount by which outputs could be increased without requiring extra inputs. Thus,

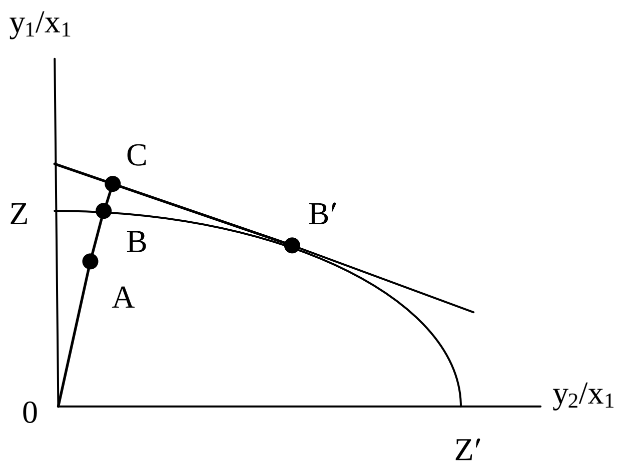
$\frac{0A}{0B}$ is the output technical efficiency of DMU A.

If price information is available, it would produce at C, and can be defined as,

$\frac{0B}{0C}$ is the output allocative efficiency of DMU A

Figure 4.3

Technical and Output Allocative Efficiencies of Output-Oriented Measurement



which has a revenue increasing interpretation (similar to the cost reducing interpretation of allocative inefficiency in input-oriented case). Furthermore, overall economics efficiency is the product of these two measurements, and can be defined as,

$\frac{0A}{0C}$ is the output overall efficiency of DMU A.

It could be written as,

$$\frac{0A}{0C} = \frac{0A}{0B}\frac{0B}{0C}$$

Therefore:

Output Overall Efficiency = Output Technical Efficiency x Output Allocative Efficiency

4.2 Modeling Methods of Comparative Performance Measurement

A simple and commonly used method for measuring the performance of an operating unit is that of performance indicators. A performance indicator is typically a ratio of some output to input pertaining to the unit being assessed.

The modeling approach to measure comparative performance attempts to arrive at a fuller understanding, for example, a model, the production process operated by the units being assessed rather than simply compute indexes of their comparative performance. There are two methods of comparative performance measurement, namely parametric and non-parametric methods.

4.2.1 Parametric Methods for Measuring Comparative Performance

Parametric methods are best illustrated in context where either a single input or alternatively a single output pertains. Thus these methods assume that they have defined a set of organizational units to be assessed and that they use a single input x to produce the outputs y_r, where r = 1…s. At this point, one of two approaches can be adopted. They could make within the model to be developed on explicit allowance for any inefficiency in production by the units being assessed, or they could make such an allowance (Ray, 2004). They make no explicit allowance in the model to be developed for any inefficiency by units being assessed, though they do not expect all of the units being assessed. Single input and multi outputs can be expressed in Equation 4.1 as following,

$$x = f(\beta, y_1, y_2 \ldots y_s) + \eta \qquad 4.1$$

where y_r, r = 1…s are the known output levels and β is a set of unknown parameters to be estimated. η is assumed to be normally distributed with mean value of zero and to be independent of the actual output levels y_r, where r = 1…s.

The method can make explicit allowance in the model to be developed for any inefficiency by units being assessed is stochastic frontier method. This method addresses two main criticisms where no explicit allowance is made in the model for any inefficiency by the unit being assessed. It estimates average rather than efficient levels of input for given outputs, and they attribute all differences between estimated and observed levels of input to inefficiency. The hypothesized version of Equation 4.2 in a stochastic frontier approach would be

$$x = f(\beta, y_1, y_2 \ldots y_s) + \nu + u \qquad 4.2$$

where ν is the random error term which is normally distributed and the term $u \geq 0$ which reflects inefficiency level.

Compared to the use of performance indicators outlined in the previous section, a parametric approach can lead to a better understanding of the production process of the units being assessed. In addition, it leads to a measure of performance rather than a multitude of performance indicators. However, the approach creates problems of its own. This can lead to a misspecified model. Another problem is that they cannot cope with multiple inputs and multiple outputs. These problems can be overcome by the non-parametric method of comparative performance measurement.

4.2.2 Non-parametric Methods for Measuring Comparative Performance

The main method in this category is Data Envelopment Analysis (DEA). In DEA, it does not hypothesize a functional form linking input and outputs. Instead, we attempt to construct a production possibility set from the observed input-output correspondences at the units being assessed. DEA is a non-parametric method that uses piecewise linear programming to calculate (rather than estimate) the efficient or best-practice frontier in a given set of decision-making units (DMUs) such as firms. The DMUs that make up the frontier envelop the less efficient firms and the relative efficiency of the firms is calculated in terms of scores on a scale of 0 to 1, with the frontier firms receiving a score of 1. DEA can calculate the allocative and technical efficiency, and the latter can be decomposed into scale, and pure technical inefficiency.

4.3 Data Envelopment Analysis (DEA)

DEA uses mathematical programming to build a non-parametric piece-wise frontier over the data. Technical efficiencies are estimated relative to this frontier[1].

An advantage of DEA is that inefficient firms are compared to actual firms rather than some statistical measure. In addition, DEA does not require specification of a cost or production function. However, the efficiency scores tend to be sensitive to the choice of input and output variables, and the method does not allow for stochastic factors and measurement errors.

Further, as more variables are included in the models, the number of firms on the frontier increases, so it is important to examine the sensitivity of the efficiency scores and rank order of the firms to model specification.

Charnes et al. (1978) proposed a model that had an input orientation and assumed constant return to scale (CRS). Subsequent papers have considered alternative sets of assumptions, such as Fare et al. (1983) & Banker et al. (1984) in which variable returns to scale (VRS) models are proposed.

4.3.1 Constant Returns to Scale (CRS) Model

Charnes et al. (1978) proposed a model that had an input orientation and assumed constant returns to scale (CRS). First assume there are data on *N* inputs and *M* outputs for each firms. For the i-th firm these are represented by the column vectors $\mathbf{x}_i$ and $\mathbf{y}_i$, respectively. The *NxI* input matrix, **X**, and *MxI* output matrix, **Y**, represent the data for all I firms (Coelli et al. 2005).

An intuitive way to introduce DEA is via the ratio form. For each firm, it obtains a measure of the ratio of all outputs over all inputs, such as $\mathbf{u}'\mathbf{y}_i/\mathbf{v}'\mathbf{x}_i$, where u is an Mx1 vector of output weights and v is a Nx1 vector of input weights. The optimal weights are obtained by solving the mathematical programming problem (Coelli et al. 2005).

---

[1] Sirasoontorn P. (March, 2005). Efficiency measures and regulation: Thai electricity generation, *Thammasat Economics Journal*, *23*(1), 38-81.

$$\begin{aligned}
&\max_{u,v}\ (\mathbf{u'y}_i/\mathbf{v'x}_i),\\
&\text{st}\quad \mathbf{u'y}_i/\mathbf{v'x}_i \le 1,\ j=1,2,\dots,I, \qquad 4.3\\
&\qquad \mathbf{u},\mathbf{v} \ge \mathbf{0}
\end{aligned}$$

From above, it points out that this model involving the ratio of outputs to inputs is referred to as the input-oriented model. Equation 4.3 is referred to as the CCR (Charnes, Cooper and Rhodes) model, and provide for constant returns to scale (CRS).

This involves finding values for **u** and **v**, such that the efficiency measure for the i-th firm is maximized, subject to the constraints that all efficiency measures must be less than or equal to one. One problem with this particular ratio formulation is that it had an infinite number of solutions. To avoid the problem, one can impose the constraint $v'x = 1$, which (Coelli et al. 2005).

$$\begin{aligned}
&\max_{\mu,\nu}\ (\boldsymbol{\mu}'\mathbf{y}_i),\\
&\text{st}\quad \boldsymbol{\nu}'\mathbf{x}_i = 1\\
&\qquad \boldsymbol{\mu}'\mathbf{y}_i/\ \boldsymbol{\nu}'\mathbf{x}_i \le 0,\ j=1,2,\dots,I, \qquad 4.4\\
&\qquad \boldsymbol{\mu},\ \boldsymbol{\nu} \ge \mathbf{0(\varepsilon)}
\end{aligned}$$

where the change of notation from **u** and **v** to **μ** and **ν** is used to point out that this is a different linear programming problem. The form of the DEA model in linear programming (LP) Equation 4.2 is known as the multiplier form (Coelli et al. 2005).

Using the duality in linear programming, one can derive an equivalent envelopment form of this problem.

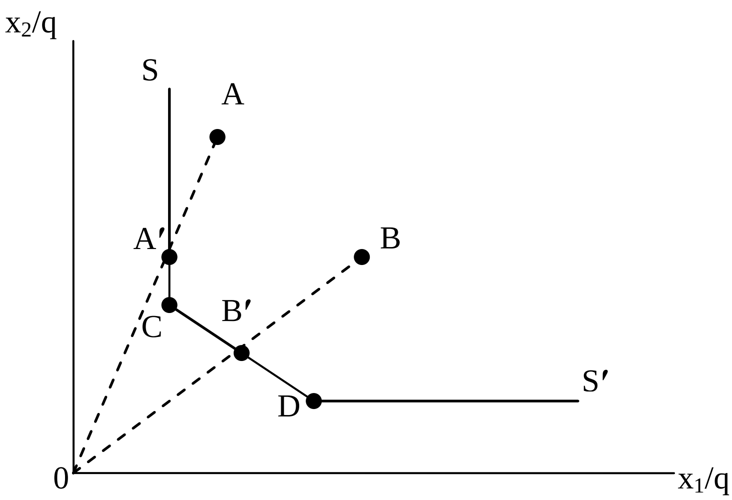
$$\begin{aligned}
&\qquad \min_{\theta,\lambda}\ \theta\\
&\text{st}\quad -\mathbf{y}_i + \mathbf{Y}\boldsymbol{\lambda} \ge \mathbf{0},\\
&\qquad \theta\mathbf{x_i} - \mathbf{X}\boldsymbol{\lambda} \ge \mathbf{0}, \qquad 4.5\\
&\qquad \boldsymbol{\lambda} \ge \mathbf{0},
\end{aligned}$$

where θ is a scalar and $\boldsymbol{\lambda}$ is a Ix1 vector of constants. This envelopment form involves fewer constraints than the multiplier form ($N+M < I+I$), and hence is generally the preferred form to solve. The value of θ obtained is the efficiency score for the i-th firm. It satisfies: $\theta \leq 1$, with a value of 1 indicating a point on the frontier and hence a technically efficient firm, according to the Farrell (1957) definition. Note that the linear programming problem must be solved I times, once for each firm in the sample. A value of θ is then obtained for each firm (Coelli et al. 2005).

The piece-wise linear form of the non-parametric frontier in DEA can cause a few difficulties in efficiency measurement. The problem arises because of the sections of the piece-wise linear frontier which run parallel to the axes (refer to the Figure 4.4) which do not occur in most parametric function.

Figure 4.4

Efficiency Measurement and Input Slacks



Source: Coelli et al. (2005).

Figure 4.3 interprets the example of efficiency measurement. Firms A and B are inefficient firms. B′ lies on a line joining points C and D. Points C and D are therefore usually referred to as the peers of point B. Point C also is the peer of point A. Farrell (1957) measure of technical efficiency gives the efficiency of firms A and B as 0A′/0A and 0B′/0B, respectively. The point A′ is an efficient point since one could reduce the amount of input x used (by the amount CA′) and still produce the same output. This is known as input slack. [2] Once one considers a case involving more inputs and/or multiple outputs, the diagrams are no longer simple, and the possibility of the related concept of output slack also occurs. Some authors argue that both the Farrell measure of technical efficiency (θ) and any non-zero input or output slacks should be reported to provide an accurate indication of technical efficiency of a firm in a DEA analysis (Coelli et al. 2005).[3]

Now it can state that, for i-th firm, the (measured) output slacks are equal to zero if $Y\lambda-y_i=0$ and the (measured) input slacks are equal to zero if $\theta x_i-X\lambda=0$ (for the given optimal values of θ and λ).

The LP in Equation 4.5 has a nice intuitive interpretation. Essentially, the problem takes the i-th firm and then seeks to radially contract the input vector, $\mathbf{x}_i$, as much as possible, while still remaining within the feasible input set. The inner-boundary of this set is a piece-wise linear isoquent. The radial contraction of the input vector, $\mathbf{x}_i$, produces a projected point, $(\mathbf{X\lambda},\mathbf{Y\lambda})$, on the surface of this technology. This projected point is a linear combination of these observed data points. The constraints in LP of Equation 4.5 ensure that this projected point cannot lie outside the feasible set (Coelli et al. 2005).

The relation between the production possibility set P and Dual variable (DLP) which is θ and $\boldsymbol{\lambda} \geq \mathbf{0}$ from Equation 4.5 can be observed. The constraints of DLP require the results of inputs and outputs $(\theta x_0, y_0)$ to belong to P, while the objective seeks the minimum θ that reduces the input vector x_0 radially to θx_0 to the possibility frontier. In DLP, it looks for an activity in P that guarantees at least the output level y_0 of DMU in all components while reducing the input vector x_0 proportionally (radially) to a value as small as possible under the consumptions that $(X\boldsymbol{\lambda}, Y\boldsymbol{\lambda})$ outperforms $(\theta x_0, y_0)$ when $\theta^* \leq 1$. With regard to this property, it can define the input excesses $\mathbf{s}^-$ and the output shortfalls $\mathbf{s}^+$ and identify them as "slack" vectors by

---

[2] Some authors use the term *input excess*.

[3] Farrell (1957) defined technical inefficiency in terms of the radial reduction in inputs that is possible.

$$\mathbf{s}^- = \theta\mathbf{x_0} - \mathbf{X\lambda},\ \ \mathbf{s}^+ = \mathbf{Y\lambda}\text{-}\mathbf{y}_0$$

with $\mathbf{s}^- \geq \mathbf{0}$, $\mathbf{s}^+ \geq \mathbf{0}$ for any feasible solution (θ, $\boldsymbol{\lambda}$) of DLP.

To discover the possible input excess and output shortfalls, it can solve the following LP using ($\boldsymbol{\lambda}$, $\mathbf{s}^-$, $\mathbf{s}^+$) as variables:

$$\begin{aligned} \max\ \ & \omega = e\mathbf{s}^- + \mathbf{es}^+ \\ \text{st.}\ \ & \mathbf{s}^- = \theta^*\mathbf{x_0} - \mathbf{X\lambda}, \\ & \mathbf{s}^+ = \mathbf{Y\lambda}\text{-}\mathbf{y}_0 \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}, \end{aligned} \qquad 4.6$$

where $\mathbf{e}$ = (1,…,1) (a vector of ones) so that e$\mathbf{s}^-$ = $\sum_{i=1}^{m} s_i^-$ and e$\mathbf{s}^-$ = $\sum_{r=1}^{m} s_r^+$. And ω = $\mathbf{w_x s^-} + \mathbf{w_y s^+}$ where $\mathbf{w_x}$ and $\mathbf{w_y}$ are positive row vectors.

The production technology associated with LP in Equation 4.5 can be defined as T= {($\mathbf{x}$,$\mathbf{y}$: $\mathbf{y} \leq \mathbf{Y\lambda}$, $\mathbf{x} \geq \mathbf{X\lambda}$, $\boldsymbol{\lambda} \geq 0$}. Fare et al. (1994) show that this technology defines a production set is closed and convex, and exhibits constant returns to scale and strong disposability (Coelli et al. 2005).

According to Figure 4.4, if point B′ is a linear combination of points C and D, where the weights in this linear combination are the λs from calculation for firm B. Firm A, technical efficiency score is 0.5 and has firm C as its peer. For example, firm A has λ_b = 0.5 and input slack 2 (IS_2) = 0.5. Point A lies on the part of the frontier that is parallel to the x_2 axis. Thus it does not represent an efficient point because it could decrease the use of the input x_2 by 0.5 units (thus producing at the point C) and still produce the same output, it can be expressed as;

0.5 x (Input 1 of A) = 0.5 x (Input 1 of firm A)
0.5 x (Input 2 of A) = 0.5 x (Input 2 of firm A) + 0.5
0.5 x (Input 3 of A) = 0.5 x (Input 3 of firm A)

Thus point A is said to be radially inefficient in input usage by a factor of 50% plus it has (non-radial) input slack of 0.5 units of x_2. The target of point A would therefore be to reduce usage of both inputs by 50% and also to reduce the use of x_2 by a further 0.5 units per unit of output).[4]

4.3.2 Variable Returns to Scale (VRS) Model

The CRS assumption is appropriate when all firms are operating at an optimal scale. However, imperfect competition, government regulations, constraints on finance, and so on, may cause a firm to be not operating at optimal scale. Banker et al. (1984) (BCC), extended the earlier work of Charnes et al. (1978) by providing for variable returns to scale (VRS). The use of the CRS specification when not all firms are operating at the optimal scale, results in measures of TE that confounded by scale efficiency (SE). The use of the VRS specification permits the calculation of TE devoid of these SE effects.

The CRS linear programming problem can be easily modified to account for VRS by adding the convexity constraint: $\mathbf{I1'\lambda} = 1$ to Equation 4.5 to provide:

$$\begin{aligned} \min_{\theta,\lambda} \quad & \theta \\ \text{st} \quad & -\mathbf{y}_i + \mathbf{Y\lambda} \geq \mathbf{0}, \\ & \theta\mathbf{x_i} - \mathbf{X\lambda} \geq \mathbf{0}, \\ & \mathbf{I1'\lambda}=1 \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \end{aligned} \qquad 4.7$$

where **I1** is an Ix1 vectors of ones. This approach forms a convex hull of interesting planes that envelop the data points more tightly than the CRS conical hull and thus provides technical efficiency scores that are greater than or equal to those obtained using the CRS model (Coelli et al. 2005).

Note that the convexity constraint ($\mathbf{I1'\lambda}$=1) essentially ensures that an efficient firm is only “benchmarked” against firms of a similar size. That is, the projected point (for that firm) on the DEA frontier is a convex combination of observed firms. This convexity restriction is not imposed in the CRS case. Hence, in a VRS DEA, a firm may be benchmark against firms that are substantially larger

[4] The output-orientation analysis will be the opposite of input- orientation analysis.

(smaller) than it. In this instance, the λ-weights sum to a value less than (greater than) one. Therefore, the returns can be considered from $\Sigma\lambda$, if $\Sigma\lambda = 1$ means the constant return to scale. If $\Sigma\lambda < 1$ means the decreasing return to scale. And if $\Sigma\lambda > 1$ means the increasing return to scale.

Technical efficiency scores obtained from the CRS model are called global technical efficiency, since the implied comparison is with firms operating at the optimal scale whereas those obtained from the VRS model are called local pure technical efficiency.

If a firm is fully efficient in both CRS and VRS specification, it is operating in most productive scale.

If a firm has a full score in VRS model but a low score in CRS model, it is operating locally efficiently but not globally efficiently due to the scale of the firm. The difference between the CRS and VRS technical efficiency scores indicates that the firm has scale inefficiency.

4.3.3 Calculation of Scale Efficiencies

Scale efficiency measures can be obtained for each firm by conducting both CRS and VRS models, and then decomposing the TE scores obtained from the CRS DEA into two components, which are scale inefficiency and “pure” technical inefficiency (or VRS TE). If there is a difference between the CRS and VRS TE scores for a particular firm, then it indicates that the firm has scale inefficiency.

Figure 4.4 illustrates scale inefficiency calculation using a one-input, one-output example. The CRS and VRS frontiers are indicated in the figure. Under CRS, the input-orientated technical inefficiency of the point P is the distance PP_C. However, under VRS, the technical inefficiency would only be PP_V. The difference between these two TE measures, P_CP_V, is due to scale inefficiency. These concepts can be expressed in ratio efficiency measures as:

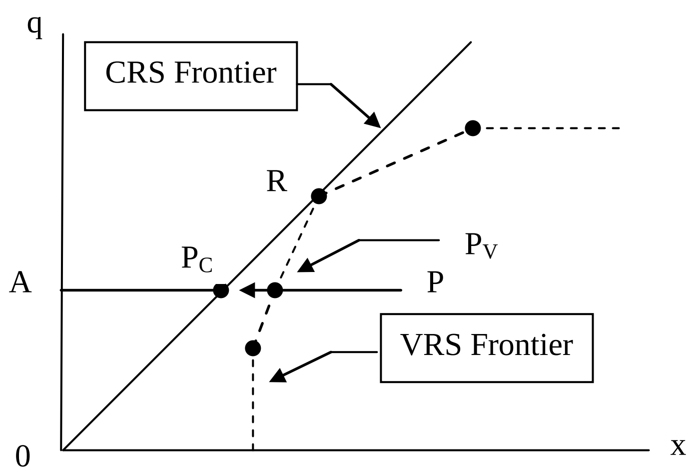
$$TE_{CRS} = AP_c/AP$$
$$TE_{VRS} = AP_V/AP \qquad 4.8$$
$$SE = AP_C/AP_V$$

where all of these measures are bounded by zero and one.

Thus, the CRS technical efficiency measure is decomposed into “pure” technical efficiency and scale efficiency. This scale efficiency measure can be roughly interpreted as the ratio of the average product of a firm operating at the point P to the average product of the point operating at the point of (technically) optimal scale (point R).

Figure 4.5

Scale Efficiency Measurement in DEA



Source: Coelli et al. (2005).

$$TE_{CRS} = TE_{VRS} \times SE \qquad 4.9$$

Because

$$AP_C/AP = (AP_V/AP)x(AP_C/AP_V) \qquad 4.10$$

4.4 Research Data Descriptions and Methodology

4.4.1 Data Descriptions and Variable Definitions

The data of BMTA are annual time series during 1989 – 2007. The technical efficiency measurement between BMTA and four selected private operators; minibus, Sahakornsong Thonburi Company, Wangsakarnkij Company, and Union Bus Service Group Company employ the annual data in 2007. This study will use Frontier 4 program under Data Envelopment Analysis (DEA) approach to measure the technical efficiency. It can provide the users to estimate the target value, lambda and slack value of inefficient firms.

The data from Sahakornsong Thonburi Company, Wangsakarnkij Company and Union Bus Service Group Company; Rangsit zone[5] are the secondary data. For minibuses, the data come from two sources. First, the number of buses is collected from the annual report of BMTA year 2007. The rests of data including of the number of employees, the fuel used, the number of passengers and the number of passengers are the average value calculating from all minibus operators in a joint conference of minibus operators that the author interviewed from one of representative of minibuses operators providing bus number 75.

This study employs two types of variables; input and output. The measurement of technical efficiency will be the case of multi input and single output.

Input

The input variables are the total number of vehicles operated by the system as the capital. They use diesel as fuel. The fuel is measured as the total annual amount of diesel used (in litres). Labor is measured as the total number of employees of the organization/firms (operators, maintenance, and administrative personnel). The numbers of employees of bus service operators mostly are the bus drivers and the conductors. The data of this input are from the annual report of BMTA. Data of number of employees in Sahakornsong Thonburi Company and Wangsakarnkij Company are obtained from their companies.

---

[5] Interview with the manager of Union Bus Service Group Company of Rangsit zone.

Output

Unlike many industries where output (e.g. perishable goods) is a clearly identifiable entity, the output of a transit firm can be quantified in various ways. A basic reason for this difference is that the ‘‘output’’ of a transit system cannot be stored for future use. Hence the study will employ two outputs, number of trips that is controlled from Deputy of Land Transport and the number of passengers which represents the real demand for bus service. They are referred as consumed output type.

The study involves the multi-input and single-output measurement. Each of the two outputs, the number of trips and that of passengers, would be selected separately to estimate with the same set of inputs.

The variables of technical efficiency measurement employed as following:

1. The number of buses (x_1) (unit: buses)
2. The number of officers (x_2) (unit: persons)
3. The amount of fuel (x_3) (unit: litres)
4. The number of trips (y_1) (unit: trips) or,
5. The number of passengers (y_2) (unit: persons)

4.4.2 Conceptual Framework

DEA is a suitable method to measure the efficiency of departments and organizations especially that of the non-profit organization (Dechpolmat, 2003). An advantage of DEA is its flexibility to estimate the efficiency of organizations by considering a lot of multi-inputs and multi-outputs that may be qualitative or quantitative variables in each time of estimation by using linear programming. The advantages of DEA are as following; firstly, DEA does not require a functional form. Secondly, this method is able to show whether firms are efficient or inefficient and demonstrate the amount of adjustment of those inputs to improve the efficiency of organizations.

BMTA is non-profit state-owned enterprise that aims to provide social service as its main responsibility. As mentioned, DEA is proper to measure the efficiency of non-profit organizations. Moreover, DEA can identify the highest efficient organization and the inefficient organizations including a suggestion for the number of input adjustments.

BMTA has loss over years causes from its inefficient operation. For the technical efficiency, BMTA can improve it but not by increasing the number of passengers because BMTA cannot control the number of passengers. It depends on the demand of commuters. But BMTA can decrease the input side to improve the technical efficiency. Therefore this study uses input-orientation DEA measurement.

DEA input-oriented models can be specified as CRS or VRS. CRS model has an assumption that firms are operating at the optimal scale. BMTA is operating in under government regulation then BMTA cannot operate at the optimal scale comparing with the other industries. Since above reason, Banker et al. extended previous model by providing for VRS model. VRS model does not have the assumption that firms are operating at optimal scale. VRS model is proper to measure efficiency for all situations whether the industries are under regulation, imperfect competition, and financial constraints and so on.

The CRS technical efficiency measurement is decomposed into pure technical efficiency and scale efficiency. If there is a different between the CRS and VRS TE scores for a particular firm, it indicates that the firm has scale inefficiency. Scale efficiency measures can be obtained for each firm by conducting both CRS and VRS models.

DEA is a non-parametric method that uses piecewise linear programming to calculate the efficient or best-practice frontier in a given set of decision-making units (DMUs) such as firms. The DMUs that make up the frontier envelop the less efficient firms and the relative efficiency of the firms is calculated in terms of scores on a scale of 0 to 1, with the frontier firms receiving a score of 1. DEA can calculate the allocative and technical efficiency, and the latter can be decomposed into scale and pure technical inefficiency.

The model with an assumption of VRS should be considered in order to eliminate the scale effect from the technical efficiency measurement by imposing the assumption of VRS into a model by adding the convexity constraint: $N1'\lambda = 1$ in to CRS model. The equations of DEA approach under CRS can be written as follows**;**

$$\begin{aligned} &\min_{\theta,\lambda} \quad \theta \\ \text{st} \quad & -\mathbf{y}_i + \mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{0}, \\ & \theta\mathbf{x}_\mathbf{i} - \mathbf{X}\boldsymbol{\lambda} \geq \mathbf{0}, \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \end{aligned} \qquad 4.11$$

To modify the CRS linear programming problem into VRS specification, will add the convexity constraint: $N1'\lambda = 1$ to Equation 4.11 as follows;

$$\begin{aligned} &\min_{\theta,\lambda} \quad \theta \\ \text{st} \quad & -\mathbf{y}_i + \mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{0}, \\ & \theta\mathbf{x}_\mathbf{i} - \mathbf{X}\boldsymbol{\lambda} \geq \mathbf{0}, \\ & \mathbf{I1'}\boldsymbol{\lambda}=1 \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \end{aligned} \qquad 4.12$$

The equations of DEA approach under CRS and VRS model that be employed in this study can be written as Equation 4.13 in the case of VRS model;

$$\begin{gathered} E_i = \min_{\theta,\lambda} \quad \theta_i \\ \text{st. } -y_{1i} + (y_{11}\lambda_1 + y_{12}\lambda_2 + y_{13}\lambda_3 + \ldots + y_{1n}\lambda_n) \geq 0 \\ (-y_{2i} + (y_{21}\lambda_1 + y_{22}\lambda_2 + y_{23}\lambda_3 + \ldots + y_{2n}\lambda_n) \geq 0) \\ \theta x_{1i} - (x_{11}\lambda_1 + x_{12}\lambda_2 + x_{13}\lambda_3 + \ldots + x_{1n}\lambda_n) \geq 0 \\ \theta x_{2i} - (x_{21}\lambda_1 + x_{22}\lambda_2 + x_{23}\lambda_3 + \ldots + x_{2n}\lambda_n) \geq 0 \\ \theta x_{3i} - (x_{31}\lambda_1 + x_{32}\lambda_2 + x_{33}\lambda_3 + \ldots + x_{3n}\lambda_n) \geq 0 \\ \lambda_1 + \lambda_2 + \lambda_3 + \ldots + \lambda_n = 1 \\ \lambda \geq 0 \end{gathered} \qquad 4.13$$

where E_i is the efficiency value of each year/firm i

y_{1i} is the number of trips of each year i of BMTA/ each firm i

x_{1i} is the number of buses of each year i of BMTA/ each firm i

x_{2i} is the number of officers of each year i of BMTA/ each firm i

x_{3i} is the amount of fuel used of each year i of BMTA/ each firm i

λ_i is the scale of year i of BMTA/ each firm i comparing with the efficiency years that lies on the frontier.

θ_i is the proportion of input used of years i of BMTA/ each firm i comparing with years that lies in the frontier.

Note: y_{2i} is the number of passengers of each year i of BMTA/ each firm i will be employed in the second technical efficiency measurement.

The linear programming problem must be solved N times, once for each firm i. The value of θ obtained if the efficiency score for the i-th firm. if θ takes a value of 1, indicating a point on the frontier which means a technically efficient firm or a best practice firm, this means that a particular firms has zero technical inefficiency. If θ is less than unity, a firm is operating or the production frontier which means that the particular firms is technically inefficient and can potentially reduce input by a factor of (1- θ) while holding output constant by adopting the behaviors of best practice firms.