CHAPTER III

RESEARCH METHODOLOGY

The study was consisted of two stages. The first stage is to measure hospital efficiency and total factor productivity (TFP) index before and after universal coverage policy with DEA. The second stage is to identify the determinants of hospitals efficiency with Tobit regression analysis.

3.1 Measurement of hospital efficiency and TFP index with Data Envelopment Analysis (DEA)

Data Envelopment Analysis (DEA) is a relatively new "data oriented" approach for evaluating the performance of a set of peer entities called Decision Making Units (DMUs) which convert multiple inputs into multiple outputs. The definition of a DMU is generic and flexible. Recent years have seen a great variety of applications of DEA for use in evaluating the performances of many different kinds of entities engaged in many different activities in many different contexts in many different countries. These DEA applications have used DMUs of various forms to evaluate the performance of entities, such as hospitals, US Air Force wings, universities, cities, courts, business firms, and others, including the performance of countries, regions, etc. Because it requires very few assumptions, DEA has also opened up possibilities for use in cases which have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in DMUs.

As pointed out in Cooper, Seiford and Tone (2000), DEA has also been used to supply new insights into activities (and entities) that have previously been evaluated by other methods. For instance, studies of benchmarking practices with DEA have identified numerous sources of inefficiency in some of the most profitable firms - firms that had served as benchmarks by reference to this (profitability) criterion – and this has provided a vehicle for identifying better benchmarks in many applied studies. Because of these possibilities, DEA studies of the efficiency of different legal organization forms such as "stock" vs. "mutual" insurance companies have shown that previous studies have fallen short in their attempts to evaluate the potentials of these different forms of organizations. Similarly, a use of DEA has suggested reconsideration of previous studies of the efficiency with which pre- and post-merger activities have been conducted in banks that were studied by DEA.

Since DEA in its present form was first introduced in 1978, researchers in a number of fields have quickly recognized that it is an excellent and easily used methodology for modeling operational processes for performance evaluations. This has been accompanied by other developments. For instance, Zhu (2002) provides a number of DEA spreadsheet models that can be used in performance evaluation and benchmarking. DEA's empirical orientation and the absence of a need for the numerous *a priori* assumptions that accompany other approaches (such as standard forms of statistical regression analysis) have resulted in its use in a number of studies involving efficient frontier estimation in the public and nonprofit sector, in the regulated sector, and in the private sector. See, for instance, the use of DEA to guide removal of the Diet and other government agencies from Tokyo to locate a new capital in Japan, as described in Takamura and Tone (2003).

In their originating study, Charnes, Cooper, and Rhodes (1978) described DEA as a 'mathematical programming model applied to observational data that provides a new way of obtaining empirical estimates of relations - such as the production functions and/or efficient production possibility surfaces – that are cornerstones of modern economies'.

Formally, DEA is a methodology directed to frontiers rather than central tendencies. Instead of trying to fit a regression plane through the *center* of the data as in statistical regression, for example, one 'floats' a piecewise linear surface to rest on top of the observations. Because of this perspective, DEA proves particularly adept at uncovering relationships that remain hidden from other methodologies. For instance, consider what one wants to mean by "efficiency", or more generally, what one wants to mean by saying that one DMU is more efficient than another DMU. This is accomplished in a straightforward manner by DEA without requiring explicitly formulated assumptions and variations with various types of models such as in linear and nonlinear regression models.

Relative efficiency in DEA accords with the following definition, which has the advantage of avoiding the need for assigning a priori measures of relative importance to any input or output, Definition 1.1 (Efficiency – Extended Pareto - Koopmans Definition): Full (100%) efficiency is attained by any DMU if and only if none of its inputs or outputs can be improved without worsening some of its other inputs or outputs.

In most management or social science applications the theoretically possible levels of efficiency will not be known. The preceding definition is therefore replaced by emphasizing its uses with only the information that is empirically available as in the following definition:

Definition 1.2 (Relative Efficiency): A DMU is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs.

Notice that this definition avoids the need for recourse to prices or other assumptions of weights which are supposed to reflect the relative importance of the different inputs or outputs. It also avoids the need for explicitly specifying the formal relations that are supposed to exist between inputs and outputs. This basic kind of efficiency, referred to as "technical efficiency" in economics can, however, be extended to other kinds of efficiency when data such as prices, unit costs, etc., are available for use in DEA.

In this study hospital efficiency is assessed using a linear programming based method, namely Data Envelopment Analysis. The method is nonparametric in nature, with no predetermined assumptions required about the functional relationship between the resource use by the hospitals and their outcomes. The method was initially proposed by Farrell and was later given in operational form by Charnes et al.

Nonparametric, non-stochastic cost and production frontiers are used to analyze the relative performance of the hospitals. Under this approach the observed input and output data are used to construct a piece-wise linear best-practice technology. The performance of each observation relative to this technology is then determined by solving a series of liear programming (LP) problems. The particular problems we solve supply input-based, radial measures of efficiency. The measures are input-based because they contract input usage while leaving outputs at their observed levels; they are radial because the input vectors are contracted along rays through the origin, thus leaving input mix unaffected. The input-based approach is chosen over the alternative output-based approach based on the assumption that cost containment (i.e., limiting input use) is a primary goal of hospital administrators and policy makers.

This particular approach is adopted for several reasons. First, it offers a convenient method for measuring operating performance that is free of monetary factors, but still has a straightforward cost interpretation. Second, the approach easily accommodates multiple outputs (hospitals offer an array of services), which is more difficult to do in an econometric setting. Third, it requires neither the assumption of a particular functional form for technology nor assumptions regarding how inefficiency is distributed, assumptions which may not be warranted. Fourth, the method does not assume that hospitals are profit-maximizers; it does, however, determine how well they utilize the available technology and how well they control their costs. The chief drawback to this approach is that it is non-stochastic; i.e., it does not allow for noise. This problem may not be as bad as it first seems. Because there is no specification of functional form, there is no specification error to show up as noise in the calculations. Furthermore, our inputs and outputs are measured in standard units, so measurement error is unlikely to be a very big problem.

For measurement of technical and scale efficiency

The CRS and VRS input-oriented DEA model will be used in this study. For the CRS input-oriented DEA we will use multi-stage, where we conduct a sequence of radial LP's to identify the efficient projected point.

In the first stage linear programming of CRS DEA we use *envelopment* form of linear programming problem:

$$
\begin{aligned}
&\min_{\theta,\lambda}\ \theta \\
&\text{subject to} \\
&\qquad Y\lambda - y_i \geq 0, \\
&\qquad \theta x_i - X\lambda \geq 0, \\
&\qquad \lambda \geq 0,
\end{aligned}
$$

where θ is a scalar and λ is a vector of constants. The value of θ obtained will be the efficiency score for the i-th DMU. It will satisfy $\theta \leq 1$, with a value of 1 indicating a point on the frontier and hence a technically efficient DMU, according to the Farrell (1957) definition.

The piecewise 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 piecewise linear frontier which run parallel to the axes which do not occur in most parametric functions. Some authors (Ali and Seiford 1993) have suggested the solution of a second-stage linear programming problem to move to an efficient frontier point by MAXIMISING the sum of slacks required to move from an inefficient frontier point to an efficient frontier point. This second stage linear programming problem may be defined by:

$$\min_{\theta,OS,IS} \quad -(OS + IS)$$

subject to

$$Y\lambda - y_i - OS = 0,$$
$$\theta x_i - X\lambda - IS = 0,$$
$$\lambda, OS, IS \geq 0,$$

where OS is an vector of output slacks and IS is a vector of input slacks. It identifies efficient projected points which have input and output mixes which are as similar as possible to those of the inefficient points, and that it is also invariant to units of measurement.

In conclusion, we will yield the radial technical efficiency score (TE), the weights in linear combination (λ) and non-radial input slacks.

The VRS input-oriented DEA model is used due to assumption that all hospitals not operating at optimal scale. Banker, Charnes and Cooper (1984) suggested an extension of the CRS DEA model to account for variable return to scale (VRS) situations. The use of the CRS specification when not all DMU's are operating at the optimal scale, will result in measures of technical efficiency (TE) which are confounded by scale efficiencies (SE). The use of the VRS specification will permit 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 to provide:

$$\min_{\theta,\lambda} \quad \theta$$

subject to

$$Y\lambda - y_i \geq 0,$$
$$\theta x_i - X\lambda \geq 0,$$
$$\Sigma\lambda = 1,$$
$$\lambda \geq 0,$$

This approach forms a convex hull of intersecting planes which envelope the data points tightly than the CRS conical hull and thus provides technical efficiency scores which are greater than or equal to those obtained using the CRS model.

Many studies have decomposed the TE scores obtained from the CRS DEA into two components, one due to scale inefficiency and one due to pure technical inefficiency. This may be done by conducting both a CRS and a VRS DEA upon the same data. If there is a difference in the two TE scores for a particular DMU, then this indicates that the DMU has scale inefficiency, and that the scale inefficiency can be calculated from the difference between the VRS TE score and the CRS TE score. The scale efficiency (SE) is equal to the ratio of the CRS TE to the VRS TE as below.

$$SE = TE_{CRS}/TE_{VRS}$$

One shortcoming of this measure of scale efficiency is that the value does not indicate whether the DMU is operating in an area of increasing or the decreasing returns to scale. This may be determined by running an addition DEA problem with non-increasing returns to scale (NIRS) imposed. This can be done by altering the DEA model by substituting the $\Sigma\lambda = 1$ restriction with $\Sigma\lambda \leq 1$, to provide:

$$\begin{aligned} &\min_{\theta,\lambda}\ \theta \\ &\text{subject to} \quad && Y\lambda - y_i \geq 0, \\ & && \theta x_i - X\lambda \geq 0, \\ & && \Sigma\lambda \leq 1, \\ & && \lambda \geq 0, \end{aligned}$$

The nature of the scale inefficiencies (i.e. due to increasing or decreasing returns to scale) for a particular DMU can be determined by seeing whether the NIRS TE score is equal to the VRS TE score. If they are unequal then increasing returns to scale exist for that DMU. If they are equal then decreasing returns to scale apply.

In conclusion, we yield the CRS and VRS TE, SE and nature of SE of each hospital.

For measurement of cost and allocative efficiency

Cost efficiency is measured via a two-step procedure. First, given technology and input prices, the cost-minimizing input vector for producing a hospital's observed level of output is calculated. Then cost efficiency is measured as the ratio of minimized cost to observed cost. The minimum cost of operation for a particular hospital is found by solving the following LP problem:

$$
\begin{aligned}
&\min_{\lambda, x_i*}\ w_i x_i^* \\
&\text{subject to} \\
&\qquad Y\lambda - y_i \ \geq 0, \\
&\qquad x_i^* - X\lambda \geq 0, \\
&\qquad \Sigma\lambda \qquad = 1, \\
&\qquad \lambda \qquad\ \ \geq 0
\end{aligned}
$$

where w_i is a vector of input prices for the i-th DMU and x_i^* (which is calculated by the LP) is the cost-minimising vector of input quantities for the i-th DMU, given the input prices w_i and the output levels y_i. The total cost efficiency (CE) or economic efficiency of the i-th DMU would be calculated as

$$CE = w_i\, x_i^* / w_i\, x_i.$$

That is, the ratio of minimum cost to observed cost. This measure indicates the proportion of observed cost required to produce the hospital's observed level of outputs. Alternatively, (1/CE - 1) gives the proportion by which observed cost exceeds best practice cost.

Failure to achieve cost efficiency may be due to the excessive use of all inputs or the incorrect mix of inputs. That is, cost-efficiency as defined above has two components - technical efficiency (TE) and allocative efficiency (AE). The technical efficiency of a particular hospital is defined as the maximum equi-proportionate contraction of the observed input vector that still allows the observed level of output to be produced.

Technical efficiency can be calculated by solving the following LP:

$$
\begin{aligned}
&\min_{\theta,\lambda}\ \theta \\
&\text{subject to} \\
&\qquad Y\lambda - y_i \ \geq 0, \\
&\qquad \theta x_i - X\lambda \geq 0, \\
&\qquad \Sigma\lambda \qquad = 1, \\
&\qquad \lambda \qquad\ \ \geq 0,
\end{aligned}
$$

The solution, TE = θ^* gives the proportion of observed inputs needed to produce the observed level of outputs with maximum efficiency, i.e., the input vector lies on the best practice frontier. When $\theta^* = 1$, it is not possible to produce the given level of outputs with fewer inputs; whereas $\theta^* < 1$ indicates that it is possible to produce the observed level of outputs using less of all inputs. The inputs are overused by the amount $(1/\theta^* - 1)$. Because this is a radial measure of technical efficiency, $(l/\theta^*$

- 1) also gives the percentage increase in cost that results due to technically inefficient performance relative to best practice. Note that the radial contraction given by θ^* may leave slack in the inputs; any slack is accounted for in the measure of allocative efficiency based on the argument that slack amounts to an inappropriate input mix.

Given measures of cost and technical efficiency, allocative efficiency can be calculated as:

$$AE = CE/TE$$

That is, only allocative efficiency will remain after the technical efficiency component is netted out of cost efficiency. In addition to quantifying the degree of allocative inefficiency, the source of this type of inefficiency can also be found. By comparing the technically efficient levels of inputs, which maintain the input mix of the observed data, to the cost efficient levels of the inputs calculated in equation above, one can determine which inputs are being over- or under-utilized relative to their cost minimizing levels.

In conclusion the total measures of hospital efficiency performance consists of cost efficiency scores, technical efficiency scores, allocative efficiency scores, scale efficiency scores and target projections.

For measuring Total Factor Productivity index before & after universal coverage policy

We use DEA-like linear programs and a (input- or output-based) Malmquist TFP index to measure productivity change, and to decompose this productivity change into technical change and technical efficiency change.

Fare et al (1994) specifies an output-based Malmquist productivity change index as:

$$m_o(y_{t+1}, x_{t+1}, y_t, x_t) = \left[\frac{d_o^t(x_{t+1}, y_{t+1})}{d_o^t(x_t, y_t)} \times \frac{d_o^{t+1}(x_{t+1}, y_{t+1})}{d_o^{t+1}(x_t, y_t)} \right]^{1/2}.$$

This represents the productivity of the production point (x_{t+1}, y_{t+1}) relative to the production point (x_t, y_t). A value greater than one will indicate positive TFP growth from period t to period t+1. This index is, in fact, the geometric mean of two output-based Malmquist TFP indices. One index uses period t technology and the other period t+1 technology. To calculate equation above we must calculate the four component distance functions, which will involve four LP problems (similar to those conducted in calculating Farrell technical efficiency (TE) measures).

The subscript "o" has been introduced to remind us that these are output-orientated measures. Note that input-orientated Malmquist TFP indices can also be defined in a similar way to the output-orientated measures presented here (Coelli).

3.2 Regression analysis for determinants of the hospital efficiency

The efficiency scores calculated by DEA are made with the assumption of homogenous inputs, outputs, and operating characteristics. But, each of these may vary from one hospital to another, and efficiency may be affected by factors representing hospital operating characteristics. In order to identify and evaluate the impact of idiosyncratic hospital-specific factors on efficiency, the efficiency score for each hospital is used as the dependent variable in a regression model. Independent variables representing the factors likely to impact on efficiency performance of the government-owned hospitals are as follows:

1) Hospital size (The number of beds): Although the size of a hospital is often cited as a factor influencing financial distress, but the patients are usually in favor of larger hospitals that offer more advanced technologies or nicer facilities and a small hospital often fails to take full advantage of increasing returns; expansion of its outputs will reduce its unit cost. This assumption is supported from the study results of Ozcan and Luke (1993) and Gerrier and Valdmanis (1996).

2) Occupancy rate: The occupancy rate is included as a measure of the demand for hospital services. It is measured as the cumulative daily census of patients to the cumulative number of beds maintained during the year. Keeping the beds full means that a hospital is producing a lot of output (i.e., patient days) from its available inputs (in large part, beds). So, a high occupancy rate should results in a high efficiency. This assumption is supported from the study results of Chang (1998).

3) Collective drug purchase involvement: In the past three years the Ministry of Public Health has the initiative in collectively drug purchase in terms of drug auction for each administration based region. We assume that involvement in such drug

auction might help hospitals to save the money and get more drugs and medical supplies for more patient treatment.

4) Geographic location: Location differences include differences in regulatory environments, demographic characteristics and socio-economic status. This assumption is supported from the study results of Gerrier and Valdmanis (1996).

5) Service complexity (type of hospital group): Recent research in management accounting and operations management suggests that service (operating) complexity reduces cost efficiency because the greater scope of services is being more complex and difficult to manage. It might be the case for technical efficiency. Therefore, the scope of services (or hospital type) is hypothesized to be negatively related to efficiency. This assumption is supported from the study results of Chang (1998).

6) Quality of services (level of hospital accreditation): One might expect quality to be negatively associated with technical efficiency as improving quality would likely require greater effort and a greater use of resources. The former finding is evidence that higher quality care requires an input mix that deviates from the efficient mix. The Pearson correlation coefficient between quality and a hospital's capital intensity is positive (though small at 0.11) and statistically significant, indicating that over-capitalization may be a means of improving the quality of care provided. This assumption is supported from the study results of Gerrier and Valdmanis (1996).

7) Competition: Hadley et al. (1996) suggests that health care reforms or market forces and competition that put financial pressure on hospitals can result in cost-containment and improved efficiency.

8) Proportion of patients under universal coverage scheme and civil servant medical benefit scheme: In Thailand, government-owned hospitals often provide free health-care services to universal coverage patients. These patients are much older or younger than the social security and civil servant patients and have a great need for additional staff time. In addition, reimbursement from civil servant patients is important source of revenue for public hospital. Most hospitals might make a greater effort for those patients to persuade them to come later.

This assumption is supported from the study results of Chang (1996) that the proportions of retired veteran patients are negatively and significantly associated with efficiency.

Hypothesis:

1) The hospital size is positively associated with technical efficiency.

2) The occupancy rate has a positive impact on technical efficiency.

3) Collective drug purchase involvement is positively associated with efficiency

4) Difference in geographic location is associated with technical efficiency.

5) Service (operating) complexity is negatively related to efficiency.

6) Quality of services (the level of HA) is negatively associated with efficiency.

7) Market competition is positively associated to technical efficiency.

8) The proportions of universal coverage and civil servant patients are negatively associated with technical efficiency.

3.3 Variables

3.3.1 Output and Input variables

An important step in hospital efficiency measurement is to specify outputs. Improved health status is the ultimate output of hospitals or the health system at large, but difficulties in accurately measuring improvements in health status. Hospital output is measured by an array of intermediate level output or so called 'process' output, namely hospital service that supposedly improve patients' health status. The inpatient service is used with some adjustment for case-mix differentials across hospitals.

However, quality adjustment in outputs could not be performed in this study due to non-availability of the necessary data on quality.

Buttler (1995) classifies hospital output into four broad categories: inpatient treatment, outpatient treatment, teaching and research. Due to not being formal teaching institution for public hospitals, so research output is not major one of public hospitals and their involvement in research is very minimal or non-existent.

Inputs in hospital production are classified as labor, capital and supplies. The labor input can be disaggregated into the various professional groups such as physician, nurse and other staff. In most studies, capital is proxied by the number of hospital beds. But, the data for supplies is not available.

The empirical DEA model is based on four inputs (total medical doctors, nursing staff, other personnel and beds) and nine outputs as table below.

Table 3.1 Output and Input Variables

Variables	Definition	Unit of measurement
Output		
Opv	Outpatient visit including health promotion and rehabilitation	The number of visits
Ipvrw	Inpatient visit adjusted with relative weight of DRG	the number of visits multiplied by average RW
Los	Total length of stay	The number of days
Teach	Teaching for intern and specialized resident doctors	The number of person
Input		
Bed	Total beds	The number of beds
Doctor	Medical doctors	The number of staff
Nurse	Registered and technical nurses	The number of staff
Other	Other personnel	The number of staff
Total personnel	Total personnel	The number of staff
Operation days	Working days including holiday	365 days
Price of labor	labor costs divided by total personnel	Amount of money (Baht)
Operation cost per day	Operational cost without labor & depreciation costs divided by total operational days per year (= 365 days)	Amount of money (Baht)
Price of beds	Net plant assets divided by total beds	Amount of money (Baht)

For productive efficiency analysis and comparative measurement of total factor productivity index and efficiency changes before and after universal coverage policy

Output Variables:

1) No. of total outpatient visit including No. of health promotion cases and No. of rehabilitation services cases

2) No. of inpatient visit adjusted with relative weight of DRG (inpatient visit multiplied by average RW of each hospital)

3) No. of inpatient days

4) No. of person trained (No. of intern trained and No. of specialized resident trained)

Input Variables:

1) No. of doctors

2) No. of nurses

3) No. of other personnel

4) No. of beds

For Cost efficiency analysis

Output Variables:

1) No. of total outpatient visit including No. of health promotion cases and No. of rehabilitation services cases

2) No. of inpatient visit adjusted with relative weight of DRG (inpatient visit multiplied by average RW of each hospital)

3) No. of inpatient days

4) No. of person trained (No. of intern trained and No. of specialized resident trained)

Input Variables:

1) Total personnel

2) Operation days (365 days)

3) Beds

4) Price of labor (total labor costs divided by total personnel)

5) Operational cost without labor & depreciation costs divided by total operational days (= 365 days)

6) Price of bed (Net plant assets divided by total beds)

3.3.2 Variables capturing hospital operating characteristics

The technical efficiency scores are examined using a censored Tobit regression model is used to identify factors that influence inefficiency.

In the Tobit model, for computational convenience, it is preferred to assume a censoring point at one. To this end, the DEA technical efficiency scores are transformed into inefficiency scores, left-censored at zero using the formula:

$$\text{Inefficiency scores (cte4rc)} = \left(\frac{1}{TEscore}\right)$$

The Tobit model is defined as follows:

$$cte4rc = 1/e = \beta_0 + \beta_1\, bed + \beta_2\, ocpr + \beta_3\, drugauct + \beta_4\, region1 + \beta_5\, region2 + \beta_6\, region4 + \beta_7\, htype1 + \beta_8\, htype2 + \beta_9\, htype4 + \beta_{10}\, htype5 + \beta_{11}\, ha1 + \beta_{12}\, ha2 + \beta_{13}\, ha3 + \beta_{14}\, compete + \beta_{15}\, opvucpc + \beta_{16}\, ipvucpc + \beta_{17}\, loscsmpc + u$$

where

cte4rc	: The reciprocal of technical efficiency scores from DEA model
bed	: the number of hospital beds
ocpr	: the occupancy rate of hospital
drugauct	: involvement in drug auction of own region (dummy variable)
region1	: Northern region (dummy variable)
region2	: Northeastern region (dummy variable)
region3	: Central region (omitted dummy variable)
region4	: Southern region (dummy variable)
htype1	: regional hospital (dummy variable)
htype2	: general hospital not less than 300 beds (dummy variable)
htype3	: general hospital less than 300 beds (omitted dummy one)
htype4	: community hospital more than 30 beds (dummy variable)
htype5	: community hospital not more than 30 beds (dummy variable)
ha0	: hospital not accredited (omitted dummy one)
ha1	: hospital accredited for level 1 (dummy variable)
ha2	: hospital accredited for level 2 (dummy variable)
ha3	: hospital accredited for highest level - level 3 (dummy variable)
compete	: compete with private hospital (dummy variable)
opvucpc	: percentage of proportion of universal coverage outpatient visit
Ipvucpc	: percentage of proportion of universal coverage inpatient visit
loscsmpc	: percentage of proportion of civil servant length of stay of inpatient visit

3.4 Data collection

The data was mainly collected from secondary sources from Bureau of health service system development, Department of Health Service Support, Ministry of

Public Health. Some part of data about teaching was collected from the Medical Council of Thailand.

3.5 Data analysis

The analysis consists of 2 stages. The first stage will be that all of hospital efficiency scores and total factor productivity (TFP) index are computed using data envelopment analysis programme, version 2.1 (DEAP 2.1) designed by Coelli. In the second stage of analysis, the estimated technical efficiency scores were regressed against a set of operating characteristics. These analyses of Tobit regression model were performed with STATA version 8.0.