ECONOMETRICS STRUCTURAL MODELING AND EFFICIENCY STUDY IN A HEALTH CARE ENVIRONMENT

ROSMANJAWATI BINTI ABDUL RAHMAN

UNIVERSITI SAINS MALAYSIA

2007

ECONOMETRIC STRUCTURAL MODELING AND EFFICIENCY STUDY IN A HEALTH CARE ENVIRONMENT

by

ROSMANJAWATI BINTI ABDUL RAHMAN

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

MAY 2007

ACKNOWLEDGEMENTS

Bismillahirrahmanirrahim.

Preparing this thesis seemed an impossible task for me especially throughout the period I was informed the most shocking news in my life – my second child, my beloved daughter, Nur Aleya Batrisyia, was diagnosed with Thalassemia Major.

However, as I sat through each page, still thinking about Aleya, makes me realize how weak I am compared to my beloved little girl. Therefore, I pulled myself together and tried to have the same courage that my little girl had and Alhamdulillah, I was able to complete this thesis. I am greatly indebted to her and would like to dedicate this thesis to her.

I am also greatly indebted to my beloved husband, Samsul Anuwar, and my son, Muhammad Aisar Faiz, and to my parents, because if not for their tolerance, patience, support and fullest cooperation, I would not have been able to complete this thesis. I must also add that it was the kind encouragement and inspiring guidance given by my Supervisor, Dr. Anton Abdulbasah Kamil of the School of Mathematical Sciences, Universiti Sains Malaysia that initially motivated me into doing this research and in helping me finishing this thesis. I'm also grateful to Assoc. Prof. Dr. Low Heng Chin that through out my thesis-writing period, she provided encouragement and sound advice and I would have been lost without her.

Finally, I cannot fail to express my profound gratitude to my sponsor, the Universiti Sains Malaysia and to all staff of the School of Mathematical Sciences, USM, for their support and encouragement especially to the Dean, Assoc. Prof. Dr. Ahmad Izani Md. Ismail and not forgetting former Dean, Prof. Dato' Rosihan M. Ali.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PUBLICATIONS & SEMINARS

1.1 Abdul Rahman, R. dan Anton Abdulbasah Kamil (2002). Mengkaji Kecekapan Unit Paediatrics, Obstetrics, Gynaecology dan Orthopaedics di Hospital Menggunakan Teknik Analisis Penyampulan Data. In Press. The National Conference on Management Science / Operations Research, Grand Continental, May 26 – 29, 2002, Kuala Terengganu.

1.2 Abdul Rahman, R. and Anton Abdulbasah Kamil (2005). Application of Structural Econometrics Model in the Healthcare System to Describe Relationships Among Variables in the System, in Proceedings of the 1st IMT – GT Regional Conference on Mathematics, Statistics and Their Applications, June 13 – 15, 2005, Parapat, Lake Toba North Sumatra. **2**: 339 – 345.

PEMBENTUKAN MODEL STRUKTUR EKONOMETRIK DAN KAJIAN KECEKAPAN DALAM PERSEKITARAN PENJAGAAN KESIHATAN

ABSTRAK

Kajian ini melibatkan kaedah Penyampulan Data untuk melihat kecekapan di empat buah unit dalam hospital iaitu unit kanak-kanak, sakit puan, bersalin dan unit ortopedik. Keputusan mendapati kesemua unit telah mencapai kecekapan maksimum melainkan

unit kanak-kanak berdasarkan input yang dikaji. Pembentukan model struktur ekonometrik pula digunakan untuk melihat situasi sebenar di dalam hospital dengan tumpuan kepada empat pemboleh ubah utama iaitu bilangan pesakit yang berdaftar, purata hari tinggal di hospital, kadar penggunaan katil dan bilangan pembedahan yang dilakukan. Model linear dan tak linear dibentuk dan dibandingkan dan didapati model linear adalah lebih baik untuk menerangkan hubungkait antara pemboleh ubah yang dikaji berdasarkan nilai punca min ralat kuasa dua. Seterusnya dengan kaedah kointegrasi dapat mengenalpasti beberapa vektor kointegrasi yang menerangkan hubungan jangka panjang dan pendek untuk setiap persamaan yang dikaji. Dari itu kesimpulan dapat dibuat wujud mekanisma keseimbangan dalam jangka masa yang panjang yang mengekalkan hubungan pemboleh ubah bersandar yang dikaji walaupun hubungan sebaliknya ditunjukkan untuk jangka masa yang pendek.

Adalah diharapkan kajian ini akan menjadi rujukan untuk mengkaji dengan lebih mendalam lagi mengenai sistem penjagaan kesihatan, khususnya perkhidmatan di hospital supaya perkhidmatan yang lebih cekap akan dapat dinikmati demi taraf kesihatan rakyat yang lebih sihat.

ECONOMETRIC STRUCTURAL MODELING AND EFFICIENCY STUDY IN A HEALTH CARE ENVIRONMENT

ABSTRACT

In this study we used Data Envelopment Analysis to evaluate relative efficiency of the Paediatrics, Obstetrics, Gynaecology and Orthopeadics units. We found that overall the performance of the studied units is good although there is room for improvement especially in the Paediatrics unit. We continued using the econometric structural modeling to look at the situation in the hospital. Four variables were looked into which included the number of registered patients, the mean duration of stay in hospital, the bed occupancy rate and total number of operations. By doing this, the linear and nonlinear relationships formed were compared and as a result we concluded that the linear model outperformed the nonlinear model based on the root of Mean Squared Error values. We then applied cointegration approach and ended-up with a number of cointegrating vectors for reliable equations that describe the long-run and short-run relationships. We concluded that there existed some long-run equilibrium mechanisms that caused different types of dependent variables in this study to remain in a relatively constant relationship, even though short-run divergence happened.

Over all, we hope this research will be a reference for further studies on our healthcare system especially in the provision of services in hospitals. The end result should be a more efficient and assessable system to increase the health standard and its awareness among the people.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

Generally, it is known that health care service is one of the government's social responsibilities to the citizens. The demand for quality health care will continue to rise in view of a growing and changing population, increasing consumer awareness (especially with the shift in population distribution from rural to urban areas), and the government's involvement in the industry. These trends have contributed to a greater government emphasis on the development and improvement of health care services in Malaysia, the responsibility of which lies with the Minister of Health.

The public health sector, heavily subsidized by the government, is the main provider of medical services in Malaysia. From the Health Facts 2004, documented by the Planning and Development Division, Ministry of Health Malaysia, there are a total of 129 public hospitals in Malaysia with an additional six specialist medical institutions including psychiatric hospitals and the National Heart Institute with 34, 414 beds. There has also been rapid growth in private hospitals in recent years. In 1980, there were only about 50 private hospitals with 2000 beds. Currently there are more than 200 private hospitals with nearly 10,000 beds, which account for 29 per cent of the total number of hospital beds in the country. Health is a significant component of total economic activity. The Ministry of Health (MOH) has been investing heavily in improving the capability of its health service and its underlying infrastructure. Public health funding has increased in line with the growth in economy, from RM 3.4 billion in 1996 to RM 8.9 billion in 2005 which covered an average of 8% of the national budget. These statistics show that health service is a significant component of total economic activity.

Measurement of efficiency in non market systems has attracted attention in current research. Evans (1971) stated that health care institutions are not always expected to be efficient. In contrast to assumed behaviour in the economic theory of the firm where efficiency is a corollary of profit maximisation, hospitals do not adhere to traditional neo-classical optimising behaviour, in part due to uncertainty caused by a lack of information on prices and costs. Thus, there is a commonly-held view, based on the length of waiting lists, media reports of patients being refused treatment, the cases of hospital closures, and so on, that the delivery of health care is inefficient. While this view is being debated, the health care expenditure keeps rising as shown throughout Malaysia's yearly plan. Data Envelopment Analysis (DEA) has widely been used to evaluate efficiency in health care system. Since it was first introduced by Charnes *et al*. (1978) and extended by Banker *et al*. (1984), DEA has been used by several researchers to study hospital performance. Sherman (1984), Grosskopf and Valdmanis (1987) and Sahin and Ozcan (2000) focused on evaluating technical efficiency of hospital production, which concerns the extent toward maximizing output production for a given level of resources and/or minimizing input usage for a given level of services produced.

Structural equation modeling (SEM) analysis has been used extensively in business and economic research (Medsker *et al*., 1994; Baumgartner and Hamburg, 1996; Chen and Steiner, 2000; Frazer, 2001; Koufteros and Marcoulides, 2006). Structural equation model is attractive because it enables researchers to test a wide range of hypotheses concerning the relationships among any combinations of manifest and latent variables. Thus, interest is also developing in the use of this method in health system evaluation. Efforts to construct the econometric structural model of health care system was initiated by Feldteins and Phil (1967) and continued by Yett *et al*. (1975). Since then, many problems related to health care system have been evaluated with the structural equation method. Some examples of the problems that have been explored thus far included the modeling of the labor market for registered nurses (Benham, 1971), describing doctor's demand in hospital (Morrisey and Jensen, 1990), relationships between market orientation and performance in the hospital (Raju *et al*., 2000), strategies for cutting hospital beds (Green and Nguyen, 2001) and the hospital technology and nurse staffing management decisions (Li and Benton, 2005).

Economic models are traditionally presented as linear models or as nonlinear models which are then linearized by the usual procedure around some equilibrium solution. But economic phenomena are not necessarily linear and, when they are nonlinear, the tendency to forget that the results obtained by the linear approximation are only locally valid may give rise to serious errors. It is only recently that nonlinear analysis has begun to be fairly widely adopted in economic model. Many problems of economics have been addressed with nonlinear modelling. Some examples are Chen and Steiner (2000), and Sakata and White (2001), Kanas and Yannopoulos (2001), Mancuso et al. (2003) and Venetis *et al*. (2003). We are also noticed that economic theory often suggests that certain pairs of variables should be linked by a long-run relationship although the variables may drift away from equilibrium. Econometricians have sought to examine and test for the presence of such a long-run equilibrium relationship between variables directly by testing whether such variables are cointegrated (Soderlind and Vredin, 1996; Christoffersen and Diebold, 1998; Huang, 2004; Rautava, 2004; and Paresh and Seema, 2005). The cointegration approach also has been applied in health economic such as Hansen and King (1996).

In this study we focus on one government's hospital and first we try to evaluate the relative efficiency of four units which are the units of Paediatrics, Gynaecology, Orthopaedics and Obstetrics. We are interested in knowing which units are more efficient in delivering the services and at the same time to determine the sources and amount of inefficiency and indicate the amount of input reduction or output increases necessary for efficiency. Here we use the Data Envelopment Analysis (DEA) technique. The rational for using DEA is its applicability to the multiple input-output nature of health care provision and the simplicity of the assumptions underlying the method. Furthermore we would like to model the 'situation' in this hospital to understand more how it operates with the number of patients increasing higher. We started with the monthly data of total number of doctors, total number of registered patients, the mean duration of stay in hospital, the beds occupancy rate, the mean duration of empty beds, total number of operations, total of patient days, the number of beds, the number of deaths, the number of discharged patients, the mean of occupied beds per day, the number of patients in first class wards and third class wards which are compiled by that hospital's record unit from January 1995 to September 2000.

We used common method in modeling econometric structural model and come out with our own linear structural model to describe relationships among variables in the health care environment. In our way to build the model we used various methods in econometric study such as rank and order condition to examine the identifiability of each equation in our model, the Breusch–Pagan/Godfrey (BP/G) test to check the assumption of the constant variance for each error terms over the observations and also the Hausman Specification Test (HST) to show that the simultaneity problem exist therefore the simultaneous equation method is appropriate and we used the Two-Stage Least Squares (2SLS) method to estimate the parameters.

So far there have been so many discussions about the nonlinearities assumption in economic model. But not much has been done in health economics field. So we hope that our nonlinear work will add up to references of nonlinear problem in healthcare. With this, then we continue with the same data, with the assumptions that there exist the nonlinear relationships among variables in the system and try to build the nonlinear model. Here, we used the Nonlinear Two-Stage Least Squares (NL2SLS) method to estimate the parameters in our nonlinear model and the comparison between this model and the above structural model are evaluated. In doing this we used the 'general-to-specific' (Lutkepohl, 2005) approach by introducing more explanatory variables in the first place including all the possible interaction variables. Then, we eliminate variables with the most statistically insignificant coefficients and re-estimate the model. These procedures were repeated until we obtain a model that contains only set of statistically significant coefficients and the model was estimated. Finally, we apply the cointegration approach to study whether there exists a long-run equilibrium relationship among our each four dependent variables with their respective set of variables that explain them.

1.2 Literature review.

Here we will look into some early works by other researchers that covered all four aspects of our study which are the efficiency study, the structural model and the nonlinear model in econometric study and also the cointegration approach.

Charnes *et al*. (1978) suggested a mathematical programming approach, referred to as Data Envelopment Analysis (DEA), to construct a frontier which envelopes all the observations to estimate the efficiencies of decision making units (DMUs). They introduced the Charnes, Cooper and Rhodes (CCR) model of DEA to evaluate the relative efficiency of decision making units (DMUs). Banker *et al*. (1984) subsequently introduced the Banker, Charnes and Cooper (BCC) model which separates technical efficiency and scale efficiency. Later, Banker (1984) showed how the CCR formulation can be employed to estimate most productive scale size and returns to scale and more recent developments described by Banker and Maindiratta (1988). To date, results of DEA have been compared to those traditional econometric techniques used for estimation of production functions. These include the method of the translog cost function that have been proposed by Christensen *et al*. (1973) and Brown *et al*. (1979).

Regression analysis is also used to evaluate the efficiency of one unit and to make a comparison between units. Regression analysis overcomes the difficulties of comparing single input to single output by estimating the average relationship between multiple inputs and outputs. Feldstein and Phil (1967) seminal study used regression analysis to determine that case-mix has an impact on hospital costs. Other examples of how regression analysis can be used include estimating marginal costs per patient, efficient rates of substitutions, fixed versus variable costs and whether economies of scale exist (Sherman, 1984). It also can be used to examine whether it is more efficient to build one large hospital or two smaller ones (Vitaliano, 1987). All the studies above show that regression analysis is useful in examining characteristics that impact costs but it is not very useful in determining an individual hospital's inefficiencies because measures of efficiency are developed by comparing decision making units to a sample mean of the characteristics.

Due to this constraint in regression analysis, frontier analysis has been developed and used to examine many important issues in the hospital industry. It has been used to examine the relative performance of public and nonprofit hospitals in California (Grosskopf and Valdmanis, 1987). In the study by Grosskopf and Valdmanis (1987), variation in input usage for different types of treatments or cases was allowed by specifying a vector of outputs rather than a single measure. They also claimed that the hospital is judged efficient if it is operating on the best practice production frontier that had been validated. Zuckerman *et al*. (1994) used frontier analysis to examine whether there is a relationship between efficiency and profitability and Hadley and Zuckerman (1991) looked into whether there is a difference in the efficiencies of urban and rural hospitals The advantages of the frontier analyses is that the decision making units

(DMUs) do not have to be individual hospitals but can be departments or resources within the hospital. It can determine the sources and amounts of inefficiency and indicate the amount of input reduction or output increases necessary for efficiency. Most studies of efficiency in the production of primary care to date have been using DEA rather than regression analysis.

The field of health economics is broadly the study of the allocation of resources to the delivery of health services. It has evolved from non quantitative studies to quantitative studies of single relationships in the health-care system, and work has already begun on formulating, estimating, and utilizing simultaneous equations models of the entire health-care system. There has also developed a quantitative approach to this field, concentrating on the econometric estimation of certain important relationships. Before we go further here are some early works that involved the construction of the structural econometric model. We start with an initial attempt to estimate a small (six-equation) econometric model of the U.S. health-care system by Feldstein and Phil (1967). At that time, each quantitative work in health economics was concerned with certain ratios, such as the physician–population ratio, and the Feldstein model was influenced by this approach. However, the model was developed to serve as a methodological prototype, not to provide detailed estimates of structural parameters of a complete model of the health care system. The second example of a simultaneous equations model of the health care system is the 47 equation macroeconometric model of Yett *et al*. (1975). In this model, the endogenous variables are described in terms of the institutions and manpower are explicitly included, whereas the exogenous and standardizing variables included demographic variables, economic variables, insurance variables, and health manpower variables. The basic mechanism of the model is that of demand and supply, however the model is not an equilibrium one. The estimated model has been used for various purposes, including forecasts of health services and health manpower and simulation of certain changes in a state health care system. Other studies that used structural model are Morrisey and Jensen (1990) that described doctor's demand in hospital, a study of three-equation of structural equations by Benham (1971) which described the labor market for registered nurses, and Green and Nguyen (2001) that suggested strategies for cutting hospital beds. They all started with formulating the structural model and ended with the estimated model to describe the situation under studied.

So far there have been so many discussions about the nonlinearities assumption in economic model. But not much has been done in health economics field. Chen and Steiner (2000) in their paper suggested a nonlinear simultaneous-equations model of analyst coverage, managerial ownerships and firm valuation. They tried to formulate a proper empirical model of these relationships by assuming that analyst coverage, managerial ownership and Tobin's Q are jointly determined and, therefore, should be modelled within three-equation system. Their argument for this empirical specification can be supported from a closer examination of earlier empirical research and the work also allowed us to gain additional insights into relationships between those three variables. They concluded that the model is better estimated compared with the same linear model.

Another work is by Kanas and Yannopoulos (2001). They compared the linear and nonlinear forecasts for stock return. The forecasting was done on the basis of forecast accuracy, using the Diebold and Mariano (1995) test and forecast encompassing, using the Clements and Hendry's (1999) approach. They employed an Artificial Neural Network (ANN) methodology to estimate a nonlinear model for stock returns, and followed with a nonlinear out–of-sample forecasting of a stock return from this model. Overall, the results showed that the inclusion of nonlinear terms in the relation between stock returns and fundamentals is important in out-of-sample forecasting. The conclusion was consistent with the view that the relation between stock returns and fundamentals is nonlinear.

There are many other works that suggested the nonlinear relations should be considered in the way to build a model. Among others there are Mancuso *et al*. (2003) that discussed the nonlinear aspects of capital market integration and real interest rate equalization, Venetis *et al*. (2003) that re-examined the predictability of economic activity using the yield spread through a nonlinear approach, Sakata and White (2001) looked into the S-estimation of nonlinear regression models with dependent and heterogenous observations and a nonlinear econometric analysis of capital flight by Schineller (1997).

Modeling the yield curve was one of the first applications of cointegration method and already considered by Engle and Granger (1987) in their seminal paper on cointegration. They introduced the representation and how to test for cointegration as well as an Error-Correction Model (ECM) and its estimation. Lanne (2000) developed a new test that is robust to deviations from the exact unit root assumption and applied to monthly US interest rate data from 1952:1 - 1991:2. While other researchers rely on the assumption that interest rates are $I(0)$ but he argued that this property cannot strictly be justified since nominal interest rates are bounded below by zero whereas $I(0)$ processes are unbounded. Christoffersen and Diebold (1998) in their paper show that imposing cointegration does not improve long-horizon forecast accuracy when forecasts of cointegrated variables are evaluated using the standard trace Mean Squared Error (MSE) ratio. They also found that by imposing cointegration on an estimated system helps the accuracy of long-horizon forecast relative to systems estimated in levels with no restrictions. Another work was by Soderlind and Vredin (1996). They used a macroeconomic equilibrium model to scrutinize some common procedures in applied cointegration analysis. In doing this they paid particular attention to cointegration relations between money, income, prices and interest rates. Their purpose was to test the hypothesis of money demand based on a Vector Error Correction Model. Huang (2004) studied the application of cointegration tests for long-run bilateral exchange rates. She investigated whether exchange rates are related to economic fundamentals in the long-run and find a range of relationships through cointegration analysis. She began by examining the time series properties of the data and using Johansen's cointegration method as well as Engle-Granger's ADF test to find evidence of cointegrating relationships. With the assumption of cointegration, she found the existence of a long-run relationship between the real exchange rates, commodity prices, nominal interest rate differential, output differential and inflation differential between Australia and New Zealand. She also performed the simple Monte Carlo study, and concluded that given a relatively short span of data it is possible for cointegration analysis to indicate that a long-run relationship had been found when in fact there is only a cyclical relationship.

Hansen and King (1996) applied the cointegration approach in their model of health care expenditure. Basically, they claimed that the stationarity of the data set is an important assumption underlying conventional regression analysis. They also argued that there is a possibility that the strong positive correlations observed between two variables were a result of non-stationarity in the respective time series, rather than evidence of an actual economic relationship. They examined this possibility in relation to a standard time series model of the macroeconomic demand for health care. Mjelde *et al*. (2002) applied the cointegration analysis to investigate relationship between six wholesale electricity markets in the western United States (U.S). They claimed that given electricity is not storable and prices based on fundamentals, the price difference between two regions in the west should be based on political structure of the trading regime and the capacity of the transmission grid. Furthermore, they said that inadequacies in either of these areas may affect price relationships. Because of this, without congestion, prices in the west should experience a high degree of cointegration. So, they concluded that price cointegration is a necessary condition for arbitrage among markets. With this, they finally examined market efficiency and stability over time in the western U.S. electricity market. Their research reached the conclusion that demand for electricity as measured by changes in cooling degree days (CDD) and heating degree days (HDD) appeared to be cointegrating factors. Including these CDD and HDD, have increased the number of cointegrating relationships, which increased the efficiency and stability of the system and ability to provide forecasts.

Another application of the cointegration approach was by Haigh (2000). He studied the relationship between freight cash and future prices using cointegration econometrics. In his article, he incorporated the long-run cointegrating relationships between cash and future prices in a forecasting model with several alternatives. Other works that used the cointegration approach in their studies are Paresh and Seema (2005) estimated income and price elasticity of imports for Fiji, Rautava (2004) studied about the impact of international oil prices and the real exchange rate on the Russian economy and its fiscal policy and Chaudhry *et al*. (1999) studied long-run stochastic properties of real estate assets by geographical breakdown.

1.3 Objectives of the study

Four major objectives are addressed in this study. There are:

1. To evaluate the relative efficiency of each unit of the Paediatrics, Obstetrics, Gynaecology and Orthopaedics in a hospital.
2. To formulate the Structural Model with the variables provided by the hospital's unit record and to look into the interaction term among that variables.
3. To build the nonlinear model from the structural model above.

4. To evaluate relationship between variables using cointegration approach.

This study will be the pioneer work for us to study more about our healthcare environment and can be a reference for more research work in health economic field in the future.

1.4 Organization of the thesis

The rest of this thesis is organized as follows. In Chapter 2, we consider all the methods of our study such as Data Envelopment Analysis (DEA), the process of Structural Equation Modeling (SEM) that covered Two-Stage Least Squares (2SLS) and Nonlinear Two-Stage Least Squares (NL2SLS). This chapter also discusses the theory of the unit root and cointegration. All the data analyses are performed in Chapter 3 and the results are discussed in Chapter 4. Finally, Chapter 5 gives the summary and conclusion of the thesis.

CHAPTER 2

THEORY AND METHODOLOGY

2.1 Introduction

Econometrics is a combination of economic theory, mathematical economics and statistics, but it is completely distinct from each one of these three branches of sciences (Koutsoyiannis, 1973). It is considered as the integration of economics, mathematics and statistics for the purpose of providing numerical values for the parameters of economic relationships and verifying economic theories. The most important characteristic of economic relationships is that they contain a random element, which is ignored by economic theory and mathematical economics. Econometrics has developed methods for dealing with this random component of economic relationships.

Much of the methodology of econometrics has been applied to various disciplines of studies such as in the military, manufacturing industry, accounting and health care system. In this chapter we will look into econometric theory and the methodology used in this research. We will start with our efficiency study which covers the topics such as the efficiency definition and measurement, followed by the description of the Data Envelopment Analysis (DEA) technique as applied to our data. Then we continue with Structural Econometric Model followed by the identification conditions, the Two-Stage Least Squares (2SLS) and the heteroscedasticity problem. Next we consider the nonlinear problem in Structural Econometric Model. First, we look into systems with nonlinearities in the variables and the a priori restrictions, and secondly, systems with nonlinearities only in the variables and the Nonlinear Two-Stage Least Squares (NL2SLS). We end this chapter with the theory of unit root and cointegration in econometrics.

2.2 Efficiency definition and measurements

There is an increasing concern with measuring and comparing the efficiency of organizational units such as local authority departments, schools, hospitals, shops, bank branches where there is a relatively homogenous set of units. In any organisation, efficiency is important to get the best result and profit. The usual measure of efficiency is

$$efficiency = \frac{output}{input} \tag{2.1}$$

If efficiency is measured wrongly, it will lead to a misallocation of resources. There are several statistical techniques to measure efficiency:

a) Ratio analysis

Ratio analysis examines the relationship between a single input and a single output. Ratios especially when tracked over time, can pinpoint changes in a hospital's operations. For example, a hospital can calculate the ratio 'cost per full time equivalent' which measures the cost per unit of staff. If this ratio is higher than other comparable hospitals, the hospital could have a problem with payroll such as excessive overtime or over-qualified staffing. It also can examine reasons for increases or decreases in costs. However, it is difficult with ratio analysis to incorporate multiple factors which is a problem since efficiency is multidimensional. Ratio analysis is useful in pinpointing specific areas of a hospital's operations that vary enough from the norm to warrant further investigation or track expenses over time but is usually not appropriate in measuring a hospital's overall efficiency.

b) Regression Analysis

Regression analysis overcomes the difficulties of comparing single input to single output by estimating the average relationship between multiple inputs and outputs. Examples of how regression analysis can be used include estimating marginal costs per patient, efficient rates of substitutions, fixed versus variable costs and whether economies of scale exist (Sherman, 1984). It also can be used to examine whether it is more efficient to build one large hospital or two smaller ones (Vitaliano, 1987). Feldstein and Phil's (1967) seminal study used regression analysis to determine that case-mix has an impact on hospital costs. It has been shown that regression analysis is useful in examining characteristics that impact costs but it is not very useful in determining an individual hospital's inefficiencies because measures of efficiency are developed by comparing firms to a sample mean.

c) Frontier Analysis

This technique uses multiple inputs and outputs from a sample of hospitals to develop an efficiency frontier and evaluate the efficiency of a Decision Making Unit (DMU) relative to all other DMUs in the sample. DMUs that are on the frontier are considered efficient while units below the frontier are considered less efficient with the distance from the frontier interpreted as the measure of inefficiency. Frontier analysis evaluates how efficient a DMU is in either producing the maximum level of outputs from a given level of inputs or using the minimum level of inputs for a given level of outputs relative to all other firms in the sample. It compares an individual hospital to the "best practice set" of the sample rather than to the sample mean. It also allows different units of measure to be used for inputs and outputs and even among inputs or outputs.

This flexibility in data definition is very helpful especially when data availability is limited, which is often the case in the public sector. This flexibility also allows for different types of hospitals in different environments with different objectives and technologies to be compared.

Frontier analysis has been used to examine many important issues in the hospital industry. It has been used to examine the relative performance of public and nonprofit hospitals in California (Grosskopf and Valdmanis, 1987), to examine whether there is a relationship between efficiency and profitability (Zuckerman *et al*. 1994), and whether there is a difference in the efficiencies of urban and rural hospitals (Hadley and Zuckerman, 1991). DMUs used in frontier analysis do not have to be individual hospitals but can be departments or resources within the hospital. It can determine the sources and amounts of inefficiency and indicate the amount of input reduction or output increases necessary for efficiency.

There are two statistical methods to identify frontier in this frontier analysis which are the Stochastic Frontier Estimation (SFE) and Data Envelopment Analysis (DEA). In this study we used DEA technique to identify the inefficient unit from four selected units which are Paediatrics, Obstetrics, Gynecology and Orthopaedics.

2.3 Stochastic Frontier Analysis (SFA)

A Stochastic Frontier Analysis (SFA) is a parametric method, developed by Aigner *et al*. (1977), Battese and Corra (1977), Jondrow *et al*. (1982), and Battese and Coelli (1988). They estimated production efficiency by introducing a two-part error term in a regression model. One is an ordinary statistical noise that accounts for measurement error and the other is a disturbance term that captures inefficiency. Moreover, Battese and Coelli (1992) assume a traditional random error ($\mathbf{V_{it}}$) and a nonnegative error term ($\mathbf{U_{it}}$) representing the technical inefficiency. Here, $\mathbf{V_{it}}$ is assumed to be independent and identically distributed, *i.i.d* $N(0,\sigma_V^2)$ and captures statistical noise, measurement error, and other random events (i.e., economic situations, quakes, weather, strikes and luck) that are beyond the company's control. The non-negative error term ($\mathbf{U_{it}}$) captures the inefficiency and is assumed to be *i.i.d* as truncations at zero of the $N(\mu,\sigma_U^2)$. Also, $\mathbf{V_{it}}$ is assumed to be independent of the $\mathbf{U_{it}}$. The model may be formed as follows:

$$\mathbf{Y_{it}} = \mathbf{X_{it}}\boldsymbol{\beta} + (\mathbf{V_{it}} - \mathbf{U_{it}}) \qquad i = 1,\ldots,K;\ t = 1,\ldots,T \tag{2.2}$$

where $\mathbf{Y_{it}}$ is output of the i^{th} firm in the t^{th} time period; $\mathbf{X_{it}}$ is a $K \times 1$ vector of inputs of the i^{th} firm in the t^{th} time period; $\boldsymbol{\beta}$ is a $K \times 1$ vector of unknown parameters; $\mathbf{V_{it}}$ and $\mathbf{U_{it}}$ are assumed to have normal and half-normal distribution, respectively.

2.4 Data Envelopment Analysis (DEA)

Data Envelopment Analysis (DEA) is a nonparametric method developed by Charnes *et al*. in 1978. It is a linear programming model, assuming no random mistakes, used to measure technical efficiency. Efficient firms are those that produce a certain amount of or more outputs while spending a given amount of inputs, or use the same amount of or less inputs to produce a given amount of outputs, as compared with other firms in the test group. The DEA method gives us a tool to estimate 'relative' efficiency of a chosen entity in a given group or units and criteria.

By maximizing (minimizing) the weighted output/input ratio of each decision making unit (DMU), an efficiency frontier can be pieced together. This ratio is less than or equal to unity for any other DMU in the data set. It measures the relative distance from the piecewise linear frontier to the DMU under evaluation. This distance falls between the values of 0 and 1. It indicates the level of input should be proportionally reduced to attain efficiency. In DEA models, we evaluate n productive units, where each DMU_s takes m different inputs to produce s different outputs. The essence of DEA models in measuring the efficiency of productive unit DMU_q lies in maximizing its efficiency rate. However, this is subjected to the condition that the efficiency rate of any other unit in the population must not be greater than 1. The models must include all characteristics considered, i.e., the weights of all inputs and outputs must be greater than zero. Such a model is defined as a linear divisive programming model:

maximize $$\frac{\sum_i u_i y_{iq}}{\sum_j v_j x_{jq}} \tag{2.3}$$

subject to $$\frac{\sum_i u_i y_{ik}}{\sum_j v_j x_{jk}} \le 1; \quad k = 1,2,\ldots,n$$

$$u_i \ge \varepsilon; \quad i = 1,2,\ldots,s$$
$$v_j \le \varepsilon; \quad j = 1,2,\ldots,m$$

where: $v_j, j = 1,2,\ldots,m,$ are weights assigned to j^{th} input,
$u_i, i = 1,2,\ldots,s$, are weights assigned to i^{th} output,
$y_{ik}, i = 1,2,\ldots s; k = 1,2,\ldots,n$, are the i^{th} outputs of k unit and
$x_{jk}, j = 1,2,\ldots m; k = 1,2,\ldots n$, are the j^{th} inputs of k unit.

This model can be converted into a linear programming model and transformed into a matrix:

maximize $$\mathbf{z} = \mathbf{u}^{\mathbf{T}}\mathbf{Y}_{\mathbf{q}} \tag{2.4}$$

subject to $$\mathbf{v}^{\mathbf{T}}\mathbf{X}_{\mathbf{q}} = \mathbf{1}$$
$$\mathbf{u}^{\mathbf{T}}\mathbf{Y} - \mathbf{v}^{\mathbf{T}}\mathbf{X} \le \mathbf{0}$$

where $\mathbf{u} \ge \boldsymbol{\varepsilon}$ and $\mathbf{v} \le \boldsymbol{\varepsilon}$

Model (2.4) is often called primary Charnes, Cooper and Rhodes (CCR) model (Charnes *et al*. 1978). The dual model to this can be stated as follows:

minimize $$\mathbf{f} = \boldsymbol{\theta} - \boldsymbol{\varepsilon}(\mathbf{e}^{\mathbf{T}}\mathbf{s}^{+} + \mathbf{e}^{\mathbf{T}}\mathbf{s}^{-}) \tag{2.5}$$

subject to $$\mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} = \mathbf{Y}_{\mathbf{q}}$$
$$\mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} = \boldsymbol{\theta}\mathbf{X}_{\mathbf{q}}$$

where $\boldsymbol{\lambda}, \mathbf{s}^{+}, \mathbf{s}^{-} \ge \mathbf{0}$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, $\boldsymbol{\lambda} \geq 0$ is a vector assigned to individual productive units, $\mathbf{s}^+$ and $\mathbf{s}^-$ are vectors of additional input and output variables, $\mathbf{e}^\mathbf{T} = (1, 1, \ldots, 1)$ and $\boldsymbol{\varepsilon}$ is a constant greater than zero, which is normally pitched at 10^{-6} or 10^{-8}. In evaluating the efficiency of unit DMU_q, model (2.5) seeks a virtual unit characterized by inputs $\mathbf{X\lambda}$ and outputs $\mathbf{Y\lambda}$, which are a linear combination of inputs and outputs of other units of the population and which are better than the inputs and outputs of unit DMU_q which is being evaluated. For inputs of the virtual unit, $\mathbf{X\lambda} \leq \mathbf{X_q}$ and for outputs $\mathbf{Y\lambda} \geq \mathbf{Y_q}$, unit DMU_q is rated efficient if no virtual unit with requested traits exists or if the virtual unit is identical with the unit evaluated, i.e., $\mathbf{X\lambda} = \mathbf{X_q}$ and $\mathbf{Y\lambda} = \mathbf{Y_q}$.

If unit DMU is CCR efficient, then the value of variable $\boldsymbol{\theta}$ is zero and also the values of all additional variables $\mathbf{s}^+$ and $\mathbf{s}^-$ equal zero. Consequently, unit DMU_q is CCR efficient if the optimum value of the model (2.5) objective function equals one. Otherwise, the unit is inefficient. The optimum value of the objective function f^* marks the efficiency rate of the unit concerned. The lower the rate, the less efficient the unit is compared to the rest of the population. In inefficient units $\boldsymbol{\theta}$ is less than one. This value shows the need for a proportional reduction of inputs for unit DMU_q to become efficient. The advantage of the DEA model is that it advises how the unit evaluated should mend its behaviour to reach efficiency.

Models (2.4) and (2.5) are input–oriented - they try to find out how to improve the input characteristics of the unit concerned for it to become efficient. There are output-oriented models as well. Such models could be written as follows:

Maximize
$$\begin{aligned} & g = \Phi + \varepsilon(e^T s^+ + e^T s^-) \\ & Y\lambda - s^+ = \Phi Y_q \\ & X\lambda + s^- = X_q \\ & \lambda, s^+, s^- \geq 0 \end{aligned} \tag{2.6}$$

This model can be interpreted as follows: unit DMU_q is CCR efficient if the optimal value of the objective function in model (2.6) equals one, $g^* = 1$. If the value of the function is greater than one, the unit is inefficient. The variable Φ indicates the need for increased output to achieve efficiency. For the optimal solution to the CCR model, the values of objective functions should be inverted, i.e., $f^* = \frac{1}{g^*}$. Models (2.4), (2.5) and (2.6) assume constant returns to scale which means that a double increase in inputs leading to a double increase in outputs. However, in efficiency analysis, variable returns to scale for, instance, an increase in inputs does not lead to the increment in outputs, can also be considered. In that case, models (2.5) and (2.6) need to be rewritten to include a condition of convexity $e^T \lambda = 1$. Afterwards, they are referred to as Banker, Charnes, Cooper (BCC) models. The aim of DEA analysis is not only to determine the efficiency rate of the units reviewed, but in particular to find target values for inputs X'_q and outputs Y'_q for an inefficient unit. After reaching these values, the unit would arrive at the threshold of efficiency. Target values are calculated using:

1. Productive unit vectors:

$$X'_q = X\lambda^*$$
$$Y'_q = Y\lambda^*$$

where λ^* is the vector of optimal variable values.

2. Efficiency rate and values of additional variables s^+ and s^-:

Input-oriented CCR model: $X'_q = \theta X_q - s^-$ and $Y'_q = Y_q + s^+$.

Output-oriented CCR model: $X'_q = X_q - s^-$

$$Y'_q = \Phi Y_q + s^+$$

where θ is the efficiency rate in the input-oriented model and Φ is the efficiency rate in the output-oriented model.

DEA and SFA have one thing in common. Both yield relative efficiency ratings on a 0 (worst-practice) to 1 (best-practice) scale based on a comparison between the observed performance of individual production units and a best-practice frontier. DEA and SFA differ across three major dimensions:

1. Nonparametric vs. parametric method.

DEA employs flexible, nonparametric methods to construct the best-practice frontier and so allows the data to ‘speak for themselves’ (Bates, Baines and Whynes, 1996). In contrast, parametric methods such as SFA assume a structure for the best practice frontier and then fit a curve.

2. Deterministic vs stochastic efficiency measurement.

DEA assumes away random error and characterizes deviations from the best-practice frontier as entirely due to inefficiency. In contrast, the stochastic frontier approach treats deviations from best practice as comprising both random error (white noise) and inefficiency.

3. Technical vs. economic efficiency.

While DEA measures technical efficiency, the SFA method measures economic efficiency. Economic efficiency is a broader term than technical efficiency. It covers an optimal choice of the level and structure of inputs and outputs based on reactions to market prices. Being economically efficient means to choose a certain volume and structure of inputs and outputs in order to minimize cost or maximize profit. Economic efficiency requires both technical efficiency and efficient allocation. While technical efficiency only requires input and output data, economic efficiency requires price data as well.

2.5 The structural econometric model

The general econometric model is an algebraic, linear (in parameters) stochastic model. Assuming there are g endogenous variables $y_1, y_2, \ldots, y_g$ and k predetermined (exogenous or lagged endogenous) variables $x_1, x_2, \ldots x_k$, the general econometric model can be written

$$y_1\gamma_{11} + y_2\gamma_{21} + \cdots + y_g\gamma_{g1} + x_1\beta_{11} + x_2\beta_{21} + \cdots + x_k\beta_{k1} = \varepsilon_1$$

$$y_1\gamma_{12} + y_2\gamma_{22} + \cdots + y_g\gamma_{g2} + x_1\beta_{12} + x_2\beta_{22} + \cdots + x_k\beta_{k2} = \varepsilon_2$$

$$\vdots \qquad (2.7)$$

$$y_1\gamma_{1g} + y_2\gamma_{2g} + \cdots + y_g\gamma_{gg} + x_1\beta_{1g} + x_2\beta_{2g} + \cdots + x_k\beta_{kg} = \varepsilon_g$$

where $\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_g$ are g stochastic disturbance terms (random variables), the γ's are coefficients of endogenous variables, and the β's are coefficients of predetermined variables. The system of equations is complete if there are as many independent equations as endogenous variables. The system of equations jointly determines values of the endogenous variables in terms of values of the predetermined variables and the values taken by the stochastic disturbance terms.

The endogenous variables are those variables which are simultaneously determined by the model and which the model is designed to explain. The exogenous variables are determined outside the model but influence the model and finally the stochastic disturbance terms are random variables that are added to all equations of the model other than identities or equilibrium conditions.

Typically, each equation of the system above has an independent meaning and identity, reflecting a behavioral relation, a technological relation or some other specific relation under study. Each equation, because it represents one aspect of the structure of the system, is called a structural equation, and the set of all structural equations is called the structural form which is the initial stage in model building. The above structural equations may be written as vector–matrix notation, in which the structural form is written as

$$\underset{1\times g}{\mathbf{y}}\;\underset{g\times g}{\mathbf{\Gamma}} + \underset{1\times k}{\mathbf{x}}\;\underset{k\times g}{\mathbf{B}} = \underset{1\times g}{\boldsymbol{\varepsilon}} \tag{2.8}$$

Here $\mathbf{y}$ and $\mathbf{x}$ are row vectors of g endogenous and k predetermined variables respectively:

$$\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \ldots & y_g \end{pmatrix} \tag{2.8.1}$$

$$\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \ldots & x_k \end{pmatrix} \tag{2.8.2}$$

And $\boldsymbol{\varepsilon}$ is a row vector consisting of g additive stochastic disturbance terms, one for each equation:

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_g \end{pmatrix} \tag{2.9}$$

The matrices $\mathbf{\Gamma}$ and $\mathbf{B}$ are the matrices of g^2 and gk structural coefficients respectively:

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1g} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2g} \\ \vdots & \vdots & & \\ \gamma_{g1} & \gamma_{g2} & \cdots & \gamma_{gg} \end{pmatrix} \tag{2.10.1}$$

$$\mathbf{B} = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1g} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2g} \\ \vdots & \vdots & & \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{kg} \end{pmatrix} \tag{2.10.2}$$

representing the complete set of coefficients of endogenous and predetermined variables respectively.

From the structural form now we can write it in the reduced form as follows, postmultiplying (2.8) by the inverse of Γ yields

$$\mathbf{y\Gamma\Gamma}^{-1} + \mathbf{xBB}^{-1} = \boldsymbol{\varepsilon}\mathbf{\Gamma}^{-1} \tag{2.11.1}$$

Thus solving for $\mathbf{y}$,

$$\mathbf{y} = -\,\mathbf{xBB}^{-1} + \boldsymbol{\varepsilon}\mathbf{\Gamma}^{-1} \tag{2.11.2}$$

which also can be written as

$$\underset{1\times g}{\mathbf{y}} = \underset{1\times k}{\mathbf{x}}\;\underset{k\times g}{\mathbf{\Pi}} + \underset{1\times g}{\mathbf{u}} \tag{2.11.3}$$

in which $$\underset{k\times g}{\mathbf{\Pi}} \equiv -\underset{k\times g}{\mathbf{B}}\;\underset{g\times g}{\mathbf{\Gamma}^{-1}} \tag{2.11.4}$$

and $$\underset{1\times g}{\mathbf{u}} \equiv \underset{1\times g}{\boldsymbol{\varepsilon}}\;\underset{g\times g}{\mathbf{\Gamma}^{-1}} \tag{2.11.5}$$

In the reduced form each of the endogenous variables is expressed as a linear function of the all predetermined variables and stochastic disturbance terms in the system. The reduced form determines the probability distributions of the endogenous variables, given the predetermined variables and given the probability distributions of the stochastic disturbance terms. The important approaches to the estimation of the