CHAPTER II

LITERATURE REVIEW

2.1 Concept of hospital efficiency measurement

Modern efficiency measurement begins with Farrell (1957) who drew upon the work of Debreu (1951) and Koopmans (1951) to define a simple measure of firm efficiency which could account for multiple inputs. He proposed that the productive efficiency of a firm consists of two components: technical efficiency, which reflects the ability of a firm to obtain maximal output from a given set of inputs, and allocative efficiency, which reflects the ability of a firm to use the inputs in optimal proportions, given their respective prices. These two measures are then combined to provide a measure of total economic efficiency (overall productive efficiency.

The cost efficiency indicates the proportion of observed cost required to produce the hospital's observed level of outputs and is measured as the ratio of minimized cost to observed cost. The cost inefficiency may be due to the excessive use of all inputs or the incorrect mix of inputs. That is, cost-efficiency has two components: technical efficiency and allocative efficiency. 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. The radial contraction in technical efficiency 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. The allocative efficiency will remain after the technical efficiency component is netted out of cost efficiency. The degree of allocative inefficiency can be quantified 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. One can determine which inputs are being over- or under-utilized relative to their cost minimizing levels.

It is often argued that health care institutions are not expected to be efficient, as they do not adhere to neo-classical firm optimization behavior. However, given the vast amount of resources that go towards funding such institutions, there is a great and growing interest in examining efficiency in hospitals with the driving force for such concern being value for money.

Different methods to test efficiency are usually considered either parametric or nonparametric, where parametric methods assume a particular functional form (such as a Cobb-Douglas production function or a translog function) and non-parametric methods do not. An alternative taxonomy is that methods can be statistical or non-statistical, where statistical methods tend to make assumptions about the stochastic nature of the data (Stochastic frontiers, as opposed to deterministic, allow for statistical 'noise'). Non-statistical methods such as DEA tend to be non-parametric (and deterministic), whereas statistical methods, based on frontier regression models tend to be parametric (and stochastic). Usually the frontier models make specific assumptions about the inefficiency term in the models which tend to be very restrictive (such as half-normal or constant inefficiency over time) (Wagstaff, 1989). SCF constructs a smooth parametric frontier which may as a result have an inappropriate technology, but accounts for stochastic error, whereas DEA constructs a piecewise linear-segmented efficiency frontier based on best practice, with no assumption about the underlying technology but no scope for random error, making it more vulnerable to data errors. DEA has the advantage that it is able to manage complex production environments with multiple input and output technologies like hospitals, but being a non-statistical method it does not produce the usual diagnostic tools with which to judge the goodness-of-fit of the model specifications produced.

Thus some trade-off exists between these methods. Non-statistical approaches such as DEA have the disadvantage of assuming no statistical noise, but have the advantage of being nonparametric and requiring no assumptions about the production frontier. SCF models on the other hand have the attraction of allowing for statistical noise, but have the disadvantage of being parametric and requiring strong assumptions about the inefficiency term. In fact, they have been criticised for their potential for mixing statistical noise and inefficiency (Skinner, 1994), particularly when the random error term does not obey the normality assumption.

In both methods, however, some non-testable assumptions have to be made. In DEA one assumes no measurement error or random fluctuations whatsoever in output and in SCF one assumes a particular error distribution. SCF has the advantage over DEA in that it may allow for measurement error, but again inefficiency is identified from a non-testable assumption about the error distribution (Newhouse, 1994). Both methods may be vulnerable to measurement and misspecification error with dangers of omitting significant variables, the inclusion of irrelevant variables, the adoption of an inappropriate technology (in SCF), or the imposition of an inappropriate variable returns to scale assumption (in DEA) (Smith, 1997). The problem of endogeneity bias, where the inputs or resources may be endogenous, has been well-documented in regression based techniques and has generally been assumed to pose no problem for DEA. However efficiency estimates in DEA may be subject to the same bias if inefficient units using low levels of the endogenous resource are set tougher efficiency targets than equally inefficient units using more of the resource (Orme & Smith, 1996). Thus pitfalls relating to errors in the measurement of the inputs and outputs and errors in specification and estimation may largely affect both techniques.

There has been a rapid increase in the application of these methods to measure hospital efficiency. However, very few studies have examined whether applying different methods to the same data will affect sensitivity of efficiency rankings.

SCF and DEA models can be compared if certain assumptions are made, such as there are no allocative inefficiencies. SCF inefficiencies can then be compared directly to those obtained from DEA. Such a study has been done by Banker, Conrad and Strauss (1986) which paid particular attention to whether there were any similarities between the two approaches in ascertaining returns to scale and technical inefficiencies. The pattern of results on the two methods, though not identical, was generally similar. When scale and technical efficiencies were combined for DEA, the two methods showed broadly similar efficiency scores. However, they argued that the methods might be sensitive to outliers and possible specification, measurement and data errors which could confound comparisons. Thus the verdict still seems to be out as to the degree of convergence between efficiency scores from the different techniques and their relative merits in measuring this.

2.2 Concept of Total Factor Productivity index

A total factor productivity (TFP) index measures change in total output relative to the change in the usage of all inputs. A TFP index is preferred over partial productivity measures, such as output per unit of labor, since partial measures can provide misleading picture of performance.

The TFP index is the ratio of the output ratio to the input ratio, for the two periods. If the level of input usage in the two periods is the same, the TFP index was composed of the change in technical efficiency and technical change. In the case where the levels of input use are different, the TFP index was composed of the change in technical efficiency and technical change and the effect of a change in the scale of operations. The scale effect is itself made up of two components namely, the scale of operations and the returns to scale. If we assume constant returns to scale (CRS) then the scale effect become unity. Otherwise we need to know the magnitude of degree of production function to be able to measure the scale effect within the TFP measure.

2.3 Previous study on hospital efficiency measurement

Masiye (2007) applied the data envelopment analysis to investigate health system performance to Zambian hospitals. The data was gathered from a sample of 30 hospitals throughout Zambia. The DEA model estimates an efficiency score for each hospital. A decomposition of technical efficiency into scale and congestion is also provided. It is found that overall Zambian hospitals are operating at 67% level of efficiency, implying that significant resources are being wasted. Only 40% of hospitals were efficient in relative terms. The study further reveals that the size of hospitals is a major source of inefficiency. Input congestion is also found to be a source of hospital inefficiency. Policy attention is drawn to unsuitable hospital scale of operation and low productivity of some inputs as factors that reinforce each other to make Zambian hospitals technically inefficient at producing and delivering services. It is argued that such evidence of substantial inefficiency would undermine Zambia's prospects of achieving its health goals.

Zere et al. (2006) applied the data envelopment analysis to measure the technical efficiency of district hospitals in Namibia. All 30 public sector hospitals were included in the study. Hospital capacity utilization ratios and the data envelopment analysis (DEA) techniques were used to assess technical efficiency. The DEA model used three inputs and two outputs. Data for four financial years (1997/98 to 2000/2001) was used for the analysis. To test for the robustness of the DEA technical efficiency scores the Jackknife analysis was used. The findings suggest the presence of substantial degree of pure technical and scale inefficiency. The average technical efficiency level during the given period was less than 75%. Less than half of the hospitals included in the study were located on the technically efficient frontier. Increasing returns to scale is observed to be the predominant form of scale inefficiency. It is concluded that the existing level of pure technical and scale inefficiency of the district hospitals is considerably high and may negatively affect the government's initiatives to improve access to quality health care and scaling up of interventions that are necessary to achieve the health-related Millennium Development Goals. It is recommended that the inefficient hospitals learn from their efficient peers identified by the DEA model so as to improve the overall performance of the health system.

Rebba and Rizzi (2006) used the data envelopment analysis to measure hospital efficiency of a sample of Italian NHS hospitals. They showed how the choice of specific constraints on input and output weights (in accordance with health care policy makers' preferences) and the consideration of exogenous variables outside the control of hospital management (and linked to past policy-makers' decisions) can affect the measurement of hospital technical efficiency using the DEA. Considering these issues, the DEA method is applied to measure the efficiency of 85 (public and private) hospitals in Veneto, a Northern region of Italy. The empirical analysis allows them to verify the role of weight restrictions and of demand in measuring the efficiency of hospitals operating within a National Health Services (NHS). They found that the imposition of lower bound on the virtual weight of acute care discharges weighted by case-mix (in order to consider policy-maker objectives) reduces average hospital efficiency. Moreover, they showed that, in many cases, low efficiency scores are attributable to external factors, which are not fully controlled by the hospital manager; especially for public hospitals low total efficiency scores can be mainly explained by past policy-makers' decisions on the size of the hospitals or their role within the regional health care service. Finally, non-profit private hospitals exhibit a higher total inefficiency while both non-profit and for-profit hospitals are characterized by higher levels of scale inefficiency than public ones.

Steinmann (2004) used the data envelopment analysis to measure and compare the (in)efficiency of German and Swiss hospitals. It seeks to answer the question of whether a given bundle of hospital services can be provided with fewer resources in the German federal state of Saxony compared to Switzerland, and whether findings are robust when attempts are made to take institutional differences into account. This study is of interest from three points of views. First, contrary to most existing work, patient days are not treated as an output but as an input. Second, the usual DEA assumption of a homogeneous sample is tested and rejected for a large part of the observations. The proposed solution is to restrict DEA to comparable observations in the two countries. The finding continues to be that hospitals of Saxony have higher efficiency scores than their Swiss counterparts. The finding proves robust with regard to modifications of DEA that are motivated by differences in hospital planning in Germany and Switzerland. The conclusions are that in Germany, the hospital remuneration scheme makes patient days the primary target variable. Moreover, the fact that the observations are planned rather than actual quantities is of minor importance. In Switzerland, quality competition is enforced to some extent by patient migration, causing the number of cases to be emphasized as an objective. Both input and output quantities suggest that the hospitals of the German sample are roughly twice as large as their Swiss counterparts. At the same time, they are far more homogeneous, which is remarkable in view of the many exclusion restrictions that had to be imposed on the Swiss sample. The larger size of German hospitals gives rise to the expectation that the DEA will indicate a larger share of units exhibiting constant and decreasing returns to scale in the German subsample. The German hospitals are more efficient on average than the Swiss. This finding is reinforced when taking into account that two-thirds of the Swiss observations cannot be projected on a German reference set, indicating that the two sets are largely disjoint. In the present DEA, calculated efficiency scores depend heavily on the standard homogeneity assumption. On the other hand, they may be considered largely robust against the choice of and changes in the exchange rate. Based on the fact that patient days relative to cases treated have been a more important performance indicator for German than for Swiss policy, counting patient days among the outputs in DEA should increase German efficiency scores more than the Swiss. This prediction is confirmed.

Linna et al. (2003) measured the productive efficiency of public dental health provision across Finland. The analysis was based on data envelopment analysis (DEA) using linear programming. In addition, they investigated various factors explaining the technical and cost efficiency of public dental care using a parametric Tobit model. These analyses revealed substantial variation in productive efficiency between health centres in different municipalities. The level of cost inefficiency was generally between 20% and 30%. Good dental health of the population, high rates of unemployment and high per capita expenditure on primary care in the municipality were associated with technical and cost inefficiency. According to the results, cost efficiency would not be improved by shifting input allocation towards more auxiliary manpower in health centres. Individual efficiency scores were clearly sensitive to the choice of output specification. Changing the unit of output measurement from visit- to patient-based measures affected markedly the ranking of dental health centres. However, the set of exogenous correlates associated to inefficiency was strikingly similar for both types of output specification. More resources are needed if the coverage of public dental care is extended to all age groups. The health centre specific efficiency scores obtained in this study can be used locally to evaluate, design and implement structural changes in the production processes.

Riedel et al. (2002) investigated the evolution of efficiency and productivity in the hospital sector of an Austrian province for the time period 1994–1996. They used panel data to design non-parametric frontier models (Data Envelopment Analysis) and compared efficiency scores and time patterns of efficiency across medical fields. As health outcomes hardly can be measured in a direct way they make use of two different approaches for output measurement: In a first approach, they employ the number of case mix-adjusted discharges and of inpatient days, in a second they use credit points, which are calculated in course of the newly introduced diagnosis related group-type financing system. They calculate and compare individual efficiency scores for hospital wards as decision making units (DMU) in specified medical fields. To their knowledge the calculation of ward-specific efficiency scores has not up till now been the unit of non-parametric efficiency analysis. Their two models find different results: Model 1 with conservative output measurement calculates an average efficiency level of 96%, while model 2 with credit points for output measurement puts average efficiency at 70%. Whereas average efficiency in model 1 hardly changes and in model 2 increases modestly in the period 1994–1996, a closer look at single hospitals displays a variety of different efficiency developments over time.

Zavras et al. (2002) used DEA to evaluate efficiency and formulate policy within a Greek national primary health care network. This study provides evidence regarding the relative efficiency of primary health care centers, as well as implications for their ideal size. Utilizing DEA, a method of proper allocation of human resources by geographic district, municipality, or community was identified. Current health sector reform efforts should be planned on the basis of such findings. Furthermore, this model should be supplemented with valid demographic, socioeconomic, and epidemiological findings. Performing stratified DEA analyses (at each regional level) may become the basis for the creation of a national health care chart, matching available resources to the population and its health needs. DEA methods can help to pinpoint good and poor clinical and administrative practices and as such, to document the necessity for creation of new facilities, consolidation, or even abolishment of inefficient and costly centers. Thus, through providing on insight into the organization and the inefficiencies of the current system and by setting quantitative and qualitative benchmarks, this method could facilitate the effort towards meeting the goal of achieving equality and efficiency in health services.

Athanassopoulos et al (1999) assessed the production and cost efficiency of 98 out of 126 hospitals of the Greek nation health system. The analysis is directly concerned with the degree of utilisation of resources and the production efficiency of the general hospitals selected. For the measurement of the indices of efficiency, the internationally known method of Data Envelopment Analysis (modified to the particular characteristics of the Greek NHS) was used. The efficiency of Greek hospitals was assessed utilising two alternative conceptual models: one focusing on production and the other on cost efficiency. The results, in both cases, indicated the scope for substantial efficiency improvements. The analysis has sought to discuss the policy implications resulting from the current efficiency status of the hospitals with reference to issues of resource re-allocation and optimal scale size.

Linna (1998) investigated the development of hospital cost efficiency and productivity in Finland in 1988–1994 using a comparative application of parametric and non-parametric panel models. Stochastic cost frontier models with a time-varying inefficiency component were used as parametric methods. As non-parametric methods various DEA models were employed to calculate efficiency scores and the Malmquist productivity index. The results revealed a 3–5% annual average increase in productivity, half of which was due to improvement in cost efficiency and half due to technological change. The results by parametric and non-parametric methods compared well with respect to individual efficiency scores, time-varying efficiency and technological change. The state subsidy reform of 1993 did not seem to have any observable effects on the hospital efficiency.

Grosskopf and Valdmanis (1993) applied the data envelopment analysis to evaluate hospital performance with case-mix-adjusted outputs. They compared hospital efficiency using a multiple input-output approach in two ways: one way used a straightforward count of inpatient days and outpatient services as outputs; and the second used a case mix-adjusted count of inpatient services and outpatient care as outputs. Their results show that there was no difference when they incorporated the case-mix index, either as a weighting device or as a separate output.

Ozcan and Luke (1993) conducted a national study of the efficiency of hospitals in urban markets. Using a sample of 3,000 urban hospitals, this article examines the contributions of selected hospital characteristics to variations in hospital technical efficiencies, while it accounts for multiple products and inputs, and controls for local environmental variations. Four hospital characteristics are examined: hospital size, membership in a multi-hospital system, ownership, and payer mix (managed care contracts, percent Medicare, and percent Medicaid). Ownership and percent Medicare are consistently found to be related significantly to hospital efficiency. Within the ownership variable, government hospitals tend to be more efficient and for-profit hospitals less efficient than other hospitals. Higher percentages of Medicare payment are negatively related to efficiency. While not consistently significant across all five of the MSA size categories in which the analyses are conducted, possession of managed care contracts, membership in a multi-hospital system and size all are consistently related positively to hospital technical efficiency. These variables are also all significant when the hospitals are examined in a combined analysis. Percent Medicaid was not significant in any of the analyses.

Sexton et al (1989) applied the methodology of data envelopment analysis (DEA) to the set of Veterans Administration medical centers (VAMC) to evaluate their relative managerial efficiencies. Each VAMC was viewed as a producer of multiple outputs and a consumer of multiple inputs. DEA uses linear programming to identify resources that were underutilized and services that were inefficiently produced. Managerial strategies based on the dual variables were constructed to indicate the manner in which inefficient VAMCs may be made efficient. The analysis showed that relative inefficiency existed in about one third of the VAMCs nationwide. Elimination of this inefficiency would save the VA over $300 million annually on personnel, equipment, drugs, and supplies, without reducing the level of services provided. A subsequent analysis of co-variance revealed that VAMCs affiliated with a university were generally less efficient than those without such an affiliation. A similar finding was obtained for larger VAMCs relative to smaller medical centers. In neither case, however, should these results be construed to imply that VAMCs should terminate their university affiliations or that VAMCs should be made smaller since factors other than relative efficiency are clearly as or more important in such decisions.

Sherman (1984) suggested a new technique for identifying inefficient hospitals, Data Envelopment Analysis (DEA), is field tested by application to a group of teaching hospitals. DEA is found to provide meaningful insights into the location and nature of hospital inefficiencies as judged by the opinion of a panel of hospital experts. DEA provides insights about hospital efficiency not available from the widely used efficiency evaluation techniques of ratio analysis and econometric regression analysis. DEA is, therefore, suggested as a means to help identify and measure hospital inefficiency as a basis for directing management efforts toward increasing efficiency and reducing health care costs.

Banker et al. (1984) showed some models for estimating technical and scale inefficiencies in data envelopment analysis. In management contexts, mathematical programming is usually used to evaluate a collection of possible alternative courses of action en route to selecting one which is best. In this capacity, mathematical programming serves as a planning aid to management. Data Envelopment Analysis reverses this role and employs mathematical programming to obtain ex post facto evaluations of the relative efficiency of management accomplishments, however they may have been planned or executed. Mathematical programming is thereby extended for use as a tool for control and evaluation of past accomplishments as well as a tool to aid in planning future activities. The CCR ratio form introduced by Charnes, Cooper and Rhodes, as part of their Data Envelopment Analysis approach, comprehends both technical and scale inefficiencies via the optimal value of the ratio form, as obtained directly from the data without requiring a priori specification of weights and/or explicit delineation of assumed functional forms of relations between inputs and outputs. A separation into technical and scale efficiencies is accomplished by the methods developed in this paper without altering the latter conditions for use of DEA directly on observational data. Technical inefficiencies are identified with failures to achieve best possible output levels and/or usage of excessive amounts of inputs. Methods for identifying and correcting the magnitudes of these inefficiencies, as supplied in prior work, are illustrated. In the present paper, a new separate variable is introduced which makes it possible to determine whether operations were conducted in regions of increasing, constant or decreasing returns to scale (in multiple input and multiple output situations). The results are discussed and related not only to classical (single output) economics but also to more modern versions of economics which are identified with "contestable market theories."

Banker (1984) estimated most productive scale size using data envelopment analysis. The relation between the most productive scale size (mpss) for particular input and output mixes and returns to scale for multiple-inputs multiple-outputs situations is explicitly developed. This relation is then employed to extend the applications of Data Envelopment Analysis (DEA) introduced by Charnes, Cooper and Rhodes (CCR) to the estimation of most productive scale sizes for convex production possibility sets. It is then shown that in addition to productive inefficiencies at the actual scale size. the CCR efficiency measure also reflects any inefficiencies due to divergence from the most productive scale size. Two illustrations of the practical applications of these results to the estimation of most productive scale sizes and returns to scale for hospitals and stem-electric generation plants are also provided to emphasize the advantage of this method in examining specific segments of the efficient production surface.

Chang (1998) combined data envelopment analysis (DEA) with regression analysis to evaluate the efficiency of central government-owned hospitals over the five fiscal years between 1990 and 1994. Efficiency is first estimated using DEA with the choice of inputs and outputs being specific to hospital operations. A multiple regression model is then employed in which the efficiency score obtained from the DEA computations is used as the dependent variable. A number of hospital operating characteristics are chosen as the independent variables. The results indicate that the scope of services and proportions of retired veteran patients are negatively and significantly associated with efficiency, whereas occupancy is positively and significantly associated with efficiency. Furthermore, the results show that hospital efficiency has improved over time during the periods studied and, given the contemporary focus on concerns regarding efficiency in health care, the results provide an indication that inter-temporal efficiency gains are attainable in healthcare sector in anticipation of the implementation of National Health Insurance Programme (Act).

Ferrier G.D. and Valdmanis V. (1996) studied rural hospital performance and its correlates. The cost, technical, allocative and scale efficiencies of a sample of rural

U.S. hospitals are calculated via linear programming models. Tobit analysis is used to assess possible correlates of each of the efficiency measures. A large amount of dispersion in operating efficiency is found within our data set; the majority of the dispersion is due to technical inefficiency. The possible correlates affecting the hospital efficiency include quality of care, size, demand for services, the mix of services offered, the intensity of care provided and location and they found that allocative and scale are found to be negatively correlated with quality. The former finding is evidence that higher quality care requires an input mix that deviates from the efficient mix. Technical efficiency and size are found to have a U-shaped relationship - large and small hospitals are relatively more technically efficient than are medium-sized hospitals. The relationship between both allocative and scale efficiency and size follows an inverted-U pattern - being either too large or too small may be deleterious. The occupancy rate is a strong, positive correlate with cost, technical and scale efficiency. The relationship between occupancy and allocative efficiency is negative. The ratio of outpatients and the intensity of care provided (the number of intensive care unit days relative to all patient care days) both have positive effects on cost efficiency, while the ratio of outpatients is positively correlated with technical and scale efficiency as well. The positive relationship between the intensity of care provided and cost efficiency is unexpected. Note, however, that the result is significant at just the 10% level and that neither of the components of cost efficiency (technical and allocative efficiency) are significantly related to the intensity of care provided. The evidence of locational differences in performance across the four states is found. In general, for-profit hospitals are found to outperform not-for-profit and public hospitals. Demand characteristics, quality of care, and the mix of services offered are also found to influence performance.

A recent General Accounting Office study [GAO (1990)] finds that hospital size, occupancy rate, ownership, and regional location are all related to financial distress and closure among hospitals.

Gianfrancesco (1990) cites the number of beds, occupancy, geographic location, and ownership as factors likely to affect efficiency.