CHAPTER III
HOSPITAL PERFORMANCE AND EFFICIENCY:
ALTERNATIVE APPROACHES

3.0. Introduction:

This chapter gives a brief outline of different approaches that are used to measure hospital efficiency. We discuss basically four approaches, viz, cost accounting, ratio analysis, combined utilization and productivity (CUP) analysis and statistical production and cost functions that are used to measure hospital performance/efficiency. These approaches have varying data requirements. The first one uses the accounting information of different hospitals and analyzes the hospital service records in order to examine hospital costs and performance. It can be applied usefully to a single hospital and involves a labor-intensive detailed examination of hospital accounts, staffing pattern, and admissions. The second and third approaches are relatively simple and less data demanding. They use hospital census data on the number of beds available in the hospital, and number of inpatients and outpatients treated. On the other hand, the fourth approach requires detailed data on hospital functioning and could be used for evaluating the efficiency of many hospitals at a time. This method, based on micro economic theory, provides insight into cost issues - the relation between marginal and average cost, the degree to which hospitals exhibit economies of scale and scope. The first two approaches do not focus on these aspects.

3.1. Cost accounting method:

The accounting method requires more detailed information on the components of hospital cost. Bamum and Kutzin (1993) have grouped cost accounting studies into
two categories The first category consists of studies which use detailed cost information in a 'step down procedure' to describe aggregate costs across departments and functions. The second category makes a less detailed estimate of hospital average cost based on aggregate central ministry information or hospital statistics and accounting records.

3.1.1, Step down procedure:

Step down cost accounting is a dis-aggregate method of analyzing the cost associated with specific hospital outputs. It is based on the scrutiny of the hospital production process to enable the best assignment of costs to outputs to which they are related. Total hospital expenditure is apportioned to specific departments (cost centers), and criteria such as 'time use' are employed to distribute all costs (including overhead and cost of intermediate outputs) to final service categories. Steps involved in a step down cost analysis are given below.

The first step in the estimation process is to get a complete picture of total hospital recurrent cost. This means combining hospital line item expenditure (typical line items are salaries, drugs, other supplies, public utilities etc.) with data on resources used that do not appear in the hospital's budget or financial statement. Next step is to attribute line item costs to cost centers in specific hospital departments. The cost centers may differ across studies. Three typical cost centers used in Barnum and Kutzin (1993) are overhead, intermediate, and final costs. The services produced by overhead cost centers are consumed by other departments (cost centers) of the hospital, not directly by patients. The cost of administration, housekeeping, and maintenance are included in this. The intermediate cost-centers generate services, which are not only consumed by various departments but also by patients directly. Examples of such costs are x-ray, laboratory
tests, operation theatre etc The final cost centers provide the services, which are directly consumed by the patients, not by other departments. These include outpatients and inpatients Some studies dis-aggregate these services among different departments, e.g., medicine, surgery, orthopedics etc

However, there are data related problems associated with this step down cost analysis. First, cost data may not be directly available for individual hospitals. Second, multiple sources of budgeting, (e.g., central as opposed to district level) and assorted ways of making payments for different line items (e.g., the salaries of physicians and nurses may be paid by health ministry, other salaries may be paid by district authorities or by the hospital) make the reconstruction of actual expenditure tedious Third, cost information may be available only on aggregate basis for the hospital The advantage, however, is the need to reconstruct hospital cost data from multiple sources provides some insight into the problems of resource allocation that confront hospital managers

3.1.2. Other accounting methods:

An alternative to step down method is to use aggregate data either for individual hospital or for a group of hospitals The method relates cost and service information to time and institutional and geographical coverage Studies based on this method require less time and less detailed information in comparison to step down procedure (Barnum and Kutzin 1993)

The above procedures provide data on average cost of different services rendered by the hospitals Though the average cost data are useful for assessing hospital performance, they are not sufficient for reaching definitive conclusions regarding hospital efficiency Further, the above cost studies underlie the following assumptions
(a) The quality of service is uniform across facilities so that the cost per an equivalent unit of output could be compared

(b) The clinical composition of patients, i.e., case mix is the same at each facility

It must be emphasized that without an understanding of quality of service and case mix, the efficiency implications of variations in average cost can not be properly interpreted. For example, high average cost may reflect high quality, but low efficiency. On the other hand, low average cost may be a reflection of low quality and thereby low-efficiency. Thus, in the absence of information on quality and case mix, it is difficult to draw meaningful conclusions from cost accounting studies.

3.2. Ratio analysis:

Another method that requires less detailed data and yet provides useful information on hospital performance is the ratio analysis. Three inter-related indicators of hospital services called average length of stay (ALS), bed occupancy rate (BOR), and bed turnover rate (BTR) are used for this purpose. These indicators are called performance indicators and are computed based on data from three basic hospital censuses, namely, bed capacity (B) of the hospital, cumulative inpatient days (IP) during a specified time interval, year in our case, and admissions (A) during the same time interval. Based on this data, the performance indicators are defined as below:

(a) Bed turnover rate or flow, \( BTR = \frac{A}{B} \)

(b) Bed occupancy rate, \( BOR = \frac{\text{IP}}{B \times \text{Time interval}} \times 100 \)
(c) Average length of stay, \( \text{ALS} = \frac{\text{IP}}{\text{A}} \)

It is to be noted that bed capacity is in the denominator of two of these three indicators, implying that bed capacity is a global surrogate for inputs. As the staffing pattern, budget etc., are broadly linked to the bed strength through norms for provision of staff and budget, it is not inappropriate to use bed capacity as a surrogate of inputs. In addition, it is to be noted that these three indicators are interdependent and based on all the three hospital censuses (B, IP, and A). Knowledge of any two can give the value for the third \(^1\). Hence, for the purpose of measurement of performance, any two indicators, say, BOA and BTR are sufficient.

### 3.3. Combined utilization and productivity analysis:

Lasso (1986) argued that the use of BOR, BTR, and ALS individually do not provide adequate information and sometimes may give misleading results on hospital performance. The bed occupancy and turnover rates (BOR and BTR) may vary from country to country and between hospitals at different levels, namely primary (low), secondary (middle) and tertiary (high). Bed occupancy rate decreases as the level of hospital decreases. Further, a low BOR increases the average cost of services. Even if hospital inputs are being used with technical efficiency, low bed occupancy implies economic inefficiency. On the other hand, a high BOR does not necessarily indicate better hospital performance. A high BOR brings down the average cost but may be associated with low quality. Thus, the implication of high BOR for average cost and hospital efficiency is ambiguous without information on other service indicators.

---

\(^1\) Annual bed occupancy measures the percentage of total available beds that are occupied by patients in a year and is calculated by the formula. \( \text{BOR} = \frac{\text{IP}}{(\text{B} \times 365)} \). Annual turnover rate is calculated by the
BOR with modest ALS may reflect an efficient situation (i.e., BTR is also high) or an inefficient situation when a large proportion of beds are occupied due to high ALS. Thus, explanatory power of a single indicator becomes clearer when information on other indicators is also available and used for the purpose (Barmm and Kutzirt, 1993).

Therefore, instead of using these indicators individually for measuring hospital performance, he developed a graphical technique for using the indicators together. Implementation of CUP analysis involves three steps: First, basic data on hospital censuses, namely inpatient days, admissions, and number of available beds need to be collected. Second, performance indicators such as bed occupancy rate (BOR), bed turnover rate (BTR), and average length of stay (ALS) are computed. Third, BOR and BTR are plotted on a chart, where average bed occupancy (also referred to as utilization) per year and the average annual number of discharges per bed (i.e., BTR, also known as productivity) are shown on the x- and y-axis respectively.

The chart is then divided into four sectors (regions) using two intersecting lines drawn at the mean values of BOR and BTR, which in turn identify the mean value of ALS. Each of the regions thus obtained has the following features: Sector I (lower left) indicates relatively low levels of bed occupancy and productivity, the least desirable situation. Sector II (upper left) indicates relatively low level of bed occupancy, but high productivity and short hospital stay. Sector III (upper right) indicates high level of bed occupancy and productivity, the most desirable situation. Sector IV (lower right) indicates relatively high level of bed occupancy, but low productivity and long hospital stay.

The formula for BTR is $BTR = \frac{A}{B}$. Average length of stay is determined by $ALS = \frac{IP}{A}$. By simple manipulation we get, $ALS = \frac{(BOR \times 365)}{BTR}$. Thus, once BOR and BTR are known, we can calculate ALS easily.

The technique of using these indicators together is called combined utilization and productivity (CUP) analysis. Lasso (1986) defines BOR as utilization and BTR as the productivity of a hospital. When these...
It is to be noted that CUP analysis can give a broad picture about the performance of hospitals. This serves as an operational/benchmark tool for the hospital administrators in order to make policy decisions. The policy maker can identify the low performing hospitals on the basis of the performance indicators for further investigation on the causes of their low performance. It is obvious that hospital performance depends very much on the quality and clinical composition of cases treated. Comparison of different hospitals about their performance is therefore problematic. CUP analysis can partly mitigate this problem by grouping the hospitals into different categories. Apart from performance indicators, there are other aspects of a hospital such as multi-product nature, organizational structure, economies of scale and scope, which are of prime importance to the researchers as well as policy makers. These could not be analyzed through any of the three techniques just described. Such issues can better be addressed using multi-product cost functions.

**3.4. Statistical/econometric approaches:**

Health economists have used statistical and econometric approaches for various purposes, namely to evaluate the efficiency of hospitals, examine the determinants of hospital costs, address the issues relating to economies of scale and scope, and to find out the relationship between marginal and average cost. These approaches are generally based on micro economic theory of production and cost. In micro economic theory of production, we come across two concepts of efficiency, namely technical and economic efficiency. Two indicators are plotted together in a chart, it shows the productivity as well as the utilization levels of the hospital under study.
3.4.1. Technical and economic efficiency:

A production process is said to be technically efficient if there is no way to produce more output with the same inputs or to produce same output with fewer inputs (Varian 1987, p 9). In the context of medical care, a process is technically efficient if production inputs (labor, drugs, equipment) are combined in a way that yields the maximum feasible output (outpatient visits, hospitalizations etc). In theory of the firm terminology, it means that any point on the production possibility frontier is technically efficient. A production process will be technically efficient only in the presence of scale effects i.e., when economies of scale exists. In other words, diseconomies of scale may arise due to managerial, technological and socio-economic inefficiencies (Sherman 1981). Technical inefficiency in case of hospitals may arise due to inefficiency in the use of staff, supplies and equipment (Barnum and Kutzin 1993).

Economic efficiency extends the concept of technical efficiency to take into account the relative costs of alternative inputs as well. A production process is called economically efficient, if inputs are combined to produce given level of output at minimum cost. Wyszewianski et al (1987) apply this concept in relative terms, arguing that efficiency improvement requires reducing the total cost of inputs used to produce a given output. In other words, if the total cost of inputs used by process X to produce a given level of output is less than the cost of inputs required by another process Y to produce the same output, then the process X is said to be more economically efficient than the process Y. The optimal input combination is given by the tangency point between an isoquant and an iso-cost (price) line. Thus, economic efficiency describes a specific subset of technically efficient choices. In fact, for a given set of input prices and
output levels, there can be only one economically efficient combination of inputs, although many technically efficient alternatives might be possible.

The statistical approaches make use of the production function in order to estimate (or otherwise characterize) the input-output relationships in health care sector. The shape of the production function is usually specified a-priori, on the basis of economic theory and practical considerations of simplicity and plausibility. Often some experimentation with alternative functional forms is required to obtain a 'best' fit or a suitable match between the data and the requirements of microeconomic theory.

Similar procedure is followed for cost function estimation. Often cost function approaches are used, in lieu of a production function to deal with multiple outputs and inputs. Thus, production function is used to deal explicitly with multiple inputs (and single output), whereas cost function deals with multiple outputs and inputs, the latter considered implicitly (Sherman 1981).

For instance, let us assume that the production function is specified as

\[ Y = F(X_1, X_2, X_m) \]  

(31)

where \( X_1, X_2, X_m \) represent the quantity/value of each of \( m \) inputs and \( Y \) the corresponding output. The functional form \( F \) is assumed to be known and each output value, \( Y \), is maximal in order that equation (31) represents a production function.

Given this situation, we may distinguish the observed input amounts for some firm by indexing them as \( X_{1j}, X_{2j}, \ldots, X_{mj} \) so that \( X_{ij} \) represents the amount of \( i^{th} \) input used by firm \( j \). Using these inputs, the \( j^{th} \) firm produces the output \( Y_j \) which satisfies

\[ 0 \leq Y_j \leq Y_j, \]  

(32)
with an efficiency ratio of unity being achieved only if production process is fully efficient. Failure to achieve a ratio of unity would then represent what might be termed as 'managerial or technical inefficiency' irrespective of the prices paid for these inputs and without regard to the prices received for the output, provided that all prices are positive.

Of course, a management, which achieves \( Y_j \), but paid excessive prices for the inputs, would simultaneously be economically inefficient. This represents price inefficiency. Closely related to the concept of price inefficiency is 'allocative inefficiency.' Such allocative inefficiency can arise, even if the prices paid are minimal, when the firm does not respond to the opportunities the market allows for substituting between factors in order to achieve a given (maximal) output level \( Y^* \). In other words, the wrong (inefficient) mix of inputs is used to achieve this output.

3.4.2. Evaluation of hospital efficiency:

Regression analysis approaches to estimating input-output relationship are widely used, but these are subject to serious limitations when applied to non-profit oriented
sectors and their efficiency evaluations. Such limitations arise from the assumption that the relationship estimated by the regression techniques reflects efficient input-output relationships. This could only be true if the observations were all generated from the hospitals that are known to be efficient. Such an assumption is questionable, to say the least, in a sector like hospital services where, clearly, pure competition in the sense of market price economics may not hold.

Another set of problems which affects most of the existing regression based studies of health care organizations is, the need to aggregate multiple outputs into a single output measure. However, we will not be dealing with the studies, which convert multiple outputs to single output. Hence, we confine to statistical approaches which deal with multiple output situations such as (a) simultaneous equation models including the recursive type and (b) single equation models, which utilize flexible functional forms to deal explicitly with multiple outputs and inputs together.

The first of these alternatives relates to causal identification via the way equations are modified relative to each other. The problem of causal identification, which occurs in simultaneous equation models, has not been addressed in most of hospital services literature, except Feldstein (1967). The translog function, a representative of the flexible functional forms, has been receiving increased attention in the study of health services efficiency. We shall examine the usefulness of this and other functional forms in analyzing hospital efficiency.

Our treatment of these statistical/econometric approaches for evaluating hospital efficiency will proceed as follows. In the following section, we will review the translog function as applied to the health sector. This is our primary interest because, (a) it
explicitly allows for simultaneous evaluation of multiple outputs and inputs which is an important attribute for hospital evaluation, and (b) it does not require that specific functional form be specified a-priori which is a critical feature of hospitals since the production and cost function of hospitals are not known with any reasonable degree of precision. Next we consider the set of hospital production and cost functions with emphasis on the study by Feldstein (1967).

3.4.3. Production and cost functions:

Production functions:

We want to consider certain background materials that are necessary for understanding flexible functional forms. The easiest way is to consider the case of an ordinary Cobb-Douglas function involving one output \( Y \) and two inputs \( X_1 \) and \( X_2 \). The function has the form

\[
Y = Y^\alpha X_1^\beta X_2^{1-\beta}
\]

Where \( \alpha, \beta \geq 0 \) and \( y \) are constants and represent output elasticities. Write factor inputs \( X_1 \) and \( X_2 \) respectively. The production process exhibits increasing, constant or decreasing returns to scale when \( \alpha + \beta > = < 1 \) respectively. The function (35) can also be written as

\[
\ln Y = \ln \gamma + \alpha \ln X_1 + \beta \ln X_2
\]

Equation (36) is the usual form for fitting the Cobb-Douglas function by the method of least squares to determine the values of \( \gamma, \alpha \) and \( \beta \). Statistical tests of significance can be applied to determine, among other things, whether \( \alpha + \beta > = < 1 \) applies. The above functional form can be extended to the case of more than two inputs,
say \textit{m}, as below

\begin{equation}
\ln Y = \ln \gamma + \sum_{i=1}^{m} a_i \cdot \ln X_i, \tag{37}
\end{equation}

It is well known that the Cobb-Douglas production function implied in both the equations (36) and (37) exhibits unitary elasticity of substitution, which is considered an unrealistic property. To avoid this, other functional forms like constant elasticity of substitution (CES), variable elasticity of substitution (VES) and a host of others are proposed in the empirical literature. Empirical analysis usually involves in specifying alternative functional forms, estimating them with the data on hand and choosing the 'best' functional form, keeping the statistical and economic criteria in mind.

Christensen, Jorgenson and Lau (1973) have proposed a general functional form known as Transcendental logarithmic form, Translog for short, to represent a second order approximation of Taylor's series expansion for any production function. This functional form has become instantaneously very popular and extended to utility, cost and profit functions as well. The advantage of this flexible functional form is that the neoclassical properties of a production function can easily be imposed as parametric restrictions and their empirical validity can be tested using data. Further, translog form allows the implementation and empirical testing of a priori restrictions of additivity/separability of factor inputs. The general translog production function can be specified as below:

\begin{equation}
\ln Y = a_0 + \sum_{i=1}^{m} a_i \ln X_i + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \beta_{ij} \ln X_i \ln X_j \tag{38}
\end{equation}

It may be noted that (38) can be estimated using single equation or systems approach. The latter is necessary when we are interested in estimating the output supply and factor demand equations corresponding to the translog production function.
Advanced econometric techniques like Zellner's seemingly unrelated regression estimation (SURE) are used in this case. Production functions such as (3.8) along with neoclassical restrictions imposed on the parameters are sometimes referred to as 'structural' models. A number of empirical studies, as already listed in Chapter II, have estimated such structural models for the health sector in developed countries. Useful discussions about efficiency (technical and economic) are carried out using the estimated production function.

Researchers have worked out the expressions for returns to scale, factor demand elasticities and Allen-Uzawa partial elasticities of substitution underlying (3.8). An additional advantage of (3.8) is that the Cobb-Douglas function given in (3.7) as well as few other functional forms such as log-quadratic can be obtained as special cases by restricting the \( \beta_i \) parameters appropriately. It is clear that (3.8) deals with a single output (Y) and several inputs (X), \( m \) in number here. Unfortunately, the multi-product generalization of the production function in (3.8) is not easy to conceive or specify. However, as we shall show shortly, the single and multi-product parallels of translog cost function are relatively easier to specify.

Recursive models:

Feldstein (1967) has proposed a recursive model to represent the production behavior of hospital services in Great Britain. The model has the ability to handle both multiple outputs and inputs, albeit in somewhat a special way. The recursive model by Feldstein (1967) is as follows:

(i) \( D = B^{\alpha_1}M^{\alpha_2}\exp(p'\beta) \)

(ii) \( W = B^{\alpha_3}M^{\alpha_4}\tilde{D}^{\alpha_5} \)
(iii) \[ R = B^\alpha_6 M^\alpha_7 \exp(p'\beta) \]  

(iv) \[ N = B^\alpha_8 R^\alpha_9 \exp(p'\beta) \]  

(v) \[ H = B^\alpha_{10} R^\alpha_{11} \exp(p'\beta) \]

where D drugs and dressing expenditures, B number of beds, M number of medical staff, W patient outputs in physical units, R proportion of occupancy, N. nursing expenditures, H housekeeping expenditures and the bracketed term is the vector of case mix proportions weighted by a vector \( \beta \).

The model has the following assumptions

(a) No hospital output is constrained by exogenously determined levels of nursing (N) and housekeeping (H) expenditures Rather, these variables are determined by the hospital, based on its level of operation

(b) Output, measured in number of patients treated, depends only on number of beds (B), medical staff (M), drugs and dressings (D) and case mix

(c) Beds (B), medical staff (M) and the vector of proportion of patient case types (p) are exogenously determined

It can be seen that in the model, each equation specifies a Cobb-Douglas type of function of the inputs and an exponential component to handle case mix variations, e.g. proportion of severe vs normal cases and ambulatory- or outpatient treatments, etc

Closure inspection of (3 9) shows that the model is recursive with two sub-systems, namely equations (i) and (ii), and equations (iii), (iv) and (v) In the first sub-system, the drugs (D), which are the outputs in equation (i) becomes input to equation (ii), but do not appear anywhere else in the system The second sub-system has proportion of occupancy (R) as an output in equation (iii) and as an input to equations (iv) and (v) The two sub-
systems are tied together by common (exogenous) input factor beds (B)

Since the above model is recursive, ordinary least squares estimation was used to estimate it in the following log-linear form

\[
\ln D = \alpha_1 \ln B + \alpha_2 \ln M + \sum_{j=1}^{k} p_j B
\]

\[
\ln W = \alpha_3 \ln B + \alpha_4 \ln M + \alpha_5 \ln D
\]

\[
\ln R = \alpha_6 \ln B + \alpha_7 \ln M + \sum_{j=1}^{k} p_j \beta_j
\]

(3.10)

\[
\ln N = \alpha_8 \ln B + \alpha_9 \ln \hat{R} + \sum_{j=1}^{k} p_j P
\]

\[
\ln H = \alpha_{10} \ln B + \alpha_{11} \ln \hat{R} + \sum_{j=1}^{k} p_j \beta_j
\]

Feldstein (1967) compared the results of this recursive model with a similar set of single equation regressions, also in Cobb-Douglas form. The comparison favored the recursive model. The results also indicate, rather weakly, the presence of decreasing returns to scale in hospital services.

Cost functions:

For a variety of reasons, it is easier to study the aspect returns to scale using cost functions (Cowing, Holtman and Powers, 1983) and this in any case, is the way most of the studies in health economics have approached this topic. In addition, many hospital studies are primarily concerned with cost behavior rather than input-output relationships addressed directly by production functions. A simple cost function may be represented as
\[ C - C(Y, p) \]  \hspace{1cm} (311)

where total cost \( C \) is a function of the output vector \( Y \) and the vector of input prices \( p \).

The cost function has the ability to consider the effect of multiple outputs on total cost. On the other hand, it deals with inputs only implicitly although, provided technical efficiency can be assumed, modern theories of Shephard-Samuelson duality permit one to move between cost and production functions to remedy this deficiency. In any case, 'p' serves as a vector of parameters with (under optimising behavior) the vector of outputs \( Y \) resulting from input mixes and magnitudes selected under whatever technology (Baumol, Panzar and Willig 1982).

Shephard [reproduced in Baumol, Panzar and Willig (1982)] used equation (311) for his duality relations between cost and production functions, which inter alia makes it possible to move in a relatively easy manner between cost and production functions. Shephard lemma (see for example, Bothwell and Cooley 1978) states

\[ \frac{\partial C(Y, p)}{\partial p_j} = X_j^* \]  \hspace{1cm} (312)

where \( X_j \) represents the optimal amount of the \( j \)th input. Therefore, at any specified level of output and prices, the optimal inputs, \( X_j \) (\( j=1, 2, \ldots, n \)), are directly determinable from the cost function.

The duality theory, which relates outputs to input in the above-indicated manner, assumes that 'technical efficiency' has already been attained in both cost and production. With this assumption, it is possible to move from cost to production functions and vice versa, so that the analyst is free to use either of the two approaches.

A specific case of equation (311) is the translog cost function, which can be
represented as

\[
\ln C = \alpha_0 + \sum_{i=1}^{m} \alpha_i \ln Y_i + \sum_{j=1}^{n} \beta_j \ln p_j + \sum_{i=1}^{m} \sum_{j=1}^{m} \delta_{ij} \ln Y_i \ln Y_j + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \phi_{ij} \ln Y_i \ln p_j
\]

(3.13)

where, \( C \) total cost, \( p_j \) unit price of \( j \)th input and \( Y_i \), amount of \( i \)th output A number of studies have used the translog cost function for measuring hospital efficiency and other related aspects

In addition to the pure translog cost function given above, some researchers have addressed the topic of efficiency as well as economies of scale and scope through hybrid functional forms In these functional forms, the independent variables, particularly the price variables, enter the equation in logarithmic form The output and other variables appear in actual scale without the logarithm The translog hybrid cost function can be specified as

\[
\ln C = \alpha_0 + \sum_{i=1}^{m} \alpha_i Y_i + \sum_{j=1}^{n} \beta_j \ln p_j + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \phi_{ij} Y_i Y_j + \]

\[
\sum_{j=1}^{n} \sum_{i=1}^{m} \gamma_{jk} \ln p_j \ln p_k + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \rho_{ij} Y_j \ln p_k
\]

(3.14)

The definition of '\( Y \)', '\( p \)' and '\( C \)' remains the same as above In the following section, we give a brief review of some of the selected functional forms that are widely used in hospital cost function studies
Hospital Cost Functions:

Feldstein (1967) approached the topic 'returns to scale' from the standpoint of cost functions. He began by examining the cost per case as a function of number of beds (B), which is a measure of scale. Two simple average cost functions were estimated with number of beds (B) in linear and quadratic form separately as below:

\[ AC = a + pB \]  
\[ AC = a + \beta_1 B + \beta_2 B^2 \]

where \( AC \) is the average cost per patient, \( a \) and \( p \) are parameters to be estimated. Though the goodness of fit was poor using the above functional forms, the results indicated increasing average cost in certain range of hospital size (number of beds). Average cost regression was re-run with a case mix variable reflecting the nine specialties (e.g., pediatrics, general surgery, general medicine etc.) added.

\[ AC = a + \rho_{fr}B + \beta_2 B^2 + \sum_{i=1}^{9} \delta_i p_i \]  

where \( p_i, (i = 1, 2, \ldots, 9) \) is the proportion of case mix types in relation to all cases treated in a hospital. In addition to substantial increase in explanatory power, the case mix variable became significant in the regression. Likewise, inclusion of case flow rate variable (average number of cases treated per bed year, \( F \)) also improved the results further.

\[ AC = a + \rho_{fr}B + ZTiPi + \delta_i F^2 \]

In equation (3.18), the average cost per patient was found to decrease as hospital bed size increases implying increasing returns to scale. In contrast, as noted earlier, the
production function estimates reflected slight decreasing returns to scale for hospitals of certain size.

Different authors used similar cost functions\(^3\). These studies on hospital cost concentrated largely upon determining the shape of the average cost curve, effects of case mix on hospital cost, and optimal hospital size by using cross sectional data on various hospitals. In addition, these studies also focus on estimating the possibility of scale economies, which implied increased efficiency for larger hospitals. However, most of these studies did not pay adequate attention to the multiple output, or input price issues as discussed in the previous chapter. Although some of these studies (particularly in 70's) gave more attention towards the case mix problems, they failed to link the empirical models used to the theoretical concept of multiple-output firm and the related cost structure. Thus, their work is considered "less regorous, even rather ad-hoc at times, than might be desired, a weakness which severely limits its interpretation and usefulness for policy analysis" (Cowing et al 1983).

More recent studies on hospital cost use theoretically consistent models of hospital cost. These studies specify the total cost to be a function of multiple categories of outputs and inputs. In addition, various measures of capital stock (e.g., number of beds, number of admitting physicians etc.), and case mix are also included in these models. These studies have used either translog or hybrid functional forms in order to estimate the relationship between the cost and output. One of the early applications of translog cost function was by Cowing and Holtman (1983).

The authors used a functional form similar to that given in equation (3.13). They

\(^3\) For example, see Carr and Feldstein (1967), New house (1970), Bays (1977), Pauly (1978), Lave and Lave (1970), and Rafferty (1971).
used five outputs (emergency room care, medical surgical care, pediatric care, maternity care and other care), prices of six inputs (nursing labor, auxiliary labor, professional labor, general labor, and materials and supplies), and two measures of fixed variables (number of admitting physicians and fixed capital) in their model. The model is specified as follows:

\[
\ln C = \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln Y_i + \sum_{j=1}^{Z} \beta_j \ln p_j + \frac{1}{2} \sum_{i=1}^{Z} \sum_{j=1}^{Z} \delta_{ij} \ln Y_i \ln Y_j + \sum_{i=1}^{Z} \sum_{j=1}^{Z} \sum_{k=1}^{Z} \gamma_{ik} \ln p_{ij} \ln p_k + \sum_{i=1}^{Z} \sum_{j=1}^{Z} \rho_{ij} \ln Y_i \ln p_j + \sum_{i=1}^{Z} \psi_i \ln Y_i \ln S + \sum_{j=1}^{n} \pi_j \ln p_j \ln S + \varphi_{sd} \ln S \ln D + \varphi_s \ln S + \varphi_{ss} (\ln S)^2 + \theta_d \log D + \theta_dd (\log D)^2
\] (3.19)

where, C: variable cost, D: number of admitting physicians (assumed to be fixed in the short run), S: measure of fixed capital stock. Using Shephard lemma, the authors derived the cost share equations for all the variable inputs. Thus, the actual model has seven equations, namely six share equations and the cost function itself. By using cross-section data from 138 non-federal, short-term, general-care hospitals in USA, the above model with six equations\(^4\) was estimated by using the method of maximum likelihood.

Probably, the most frequently used functional form in recent times is the hybrid cost function originally proposed by Grannemann et al. (1986). The functional form is given below:

\(^4\) Since cost shares add-up to unity, out of the six dare equations only five are linearly independent and can be estimated. The parameters of the sixth equation can be obtained using all other parameters.
\[ C = A \prod_{i=1}^{n} p_i^{\beta_i} \exp(f(Y, D, CM, R, Z)) \]  \hspace{1cm} (3.20)

or

\[ \ln C = \ln A + \sum_{i=1}^{n} \beta_i \ln p_i + f(Y, D, CM, R, Z) \]  \hspace{1cm} (3.21)

where \( A \): vector of various factors that are assumed to affect the level of cost, but not the shape of the cost function with respect to outputs, \( P_j \): price of \( j^{th} \) production input, \( Y \): vector of hospital primary outputs (\( Y_1, Y_2, Y_3 \) are inpatient days by type, \( Y_4 \) is outpatient visits, \( Y_5 \) emergency department visits), \( D \): vector of inpatient discharges by type, \( CM \): matrix of case mix variables, \( R \): vector of various sources of revenue of the hospital, \( Z \): vector of other miscellaneous outputs produced by the hospital. The outputs enter the equation in linear, quadratic as well as cubic forms and other variables such as CM and R enter with interaction terms.

Functional forms similar to Grannemann et al (1986) have been used in Bitran and Dunlop (1989), Vita (1990), Barnum and Kutzin (1993) and various other studies.

Mixed models:

The above models are primarily used for examining the determinants of hospital costs, estimating the extent of scale and scope of hospitals. However, in studies where the primary focus is to examine the efficiency of the hospital production process, use somewhat different approach. These studies, in addition to the use of cost functions, use the production function approach in order to calculate the efficiency index. Goldman and Grossman (1983) have used this approach for estimating the inefficiency of community health centers in the U S. Similar approach was followed by Wouters (1990) in her study of cost and efficiency of a sample of 42 private and public hospitals in Nigeria.
Technical efficiency was estimated using a production function. Measures of marginal product of health workers are also obtained. The following production function was used:

\[ \ln V = a_0 + a_1 \left( \frac{\ln (IN) - 1}{\ln DRUGNUM86 + a_3 \ln HHW + a_4 \ln LHW} \right) + a_5 \ln \text{BEDSDUMMY} + a_6 \left( \frac{\ln \text{BEDS}}{\ln \text{BEDSDUMMY}} \right) \]

(3.22)

where \( \lambda \) is a parameter of the Box-Cox transformation. In the above production function, (logarithm of) the number of outpatient visits (\( V \)) is the dependent variable. Independent variables are, the number of inpatient admissions (\( IN \)), the number of patients receiving drugs (\( DRUGNUM86 \)), the proportion of high and low level health workers (\( HHW, LHW \)) and the presence of number of beds (\( \text{BEDSDUMMY} \) and \( \text{BEDS} \)).

To examine whether the cost minimization was taking place, estimates of marginal product of high and low level worker were compared with relative category of labor in private and public facilities. She found that public facilities employed too many low-level workers compared to private hospitals. In contrast, private providers were found to be using a near optimal (cost minimizing) mix of high and low level of health workers. The estimated cost function is as follows:

\[ \ln \text{RCOST} = b_0 + b_1 \ln V + b_2 \left( \frac{\ln (IN) - 1}{\ln DRUGPCT86} \right) + b_3 \ln \text{INDEX} - b_4 \ln \text{WAGHHW} + b_5 \ln \text{WAGLHW} + b_7 \ln \text{BEDSDUMMY} + \frac{b_8 \left( \frac{\ln \text{BEDS} - 1}{\ln \text{BEDSDUMMY}} \right)}{\lambda} \]

(3.23)
where RCOST recurrent expenditure of the facility, DRGPCT86 proportion of patients obtaining drugs, WAGHHW and WAGLHW wages of high and low level health workers and INDEX efficiency index defined as,

\[
\text{INDEX} = \frac{\text{MP}_{\text{LHW}}}{\text{MP}_{\text{HHW}}} \cdot \frac{\text{WAGE}_{\text{LHW}}}{\text{WAGE}_{\text{HHW}}} - 1
\]  

(3.24)

The other variables are as defined in equation (3 23) The equation (3 23) was estimated using OLS method

Thus, in this chapter, we have reviewed the alternative approaches that were used in the literature for evaluation of hospital performance and efficiency We notice that the method of combined utilization and productivity (CUP) was used for evaluating hospital performance and statistical cost functions for hospital (economic/allocative) efficiency

CUP analysis is a managerial technique that could be used for quick identification of low performing hospitals. This method does not require detailed data on the hospital production process and depends mostly on hospital census data such as inpatient days, outpatient visits, admissions, beds etc However this method does not take certain critical issues such as economies of scale and scope, marginal and average costs, elasticity of substitution etc, which are of prime importance to the economists and policy makers These issues are well handled by the statistical cost functions We therefore, attempt to use the CUP analysis for evaluating hospital performance and cost functions for efficiency in Andhra Pradesh The specific objectives of this study are listed below
3.5. **Specific objectives of the study:**

1. Evaluate the performance and efficiency of secondary level hospitals in the state of Andhra Pradesh. This is attempted through:
   (a) Combined utilization and productivity (CUP) analysis for evaluating hospital performance.
   (b) Cost functions for evaluating hospital efficiency.

2. Examining the issue of health financing in developing countries in general and AP in particular, with a view to suggest some policy measures in this regard. This is done through:
   (a) Review the methods of financing the health sector as suggested by various international agencies.
   (b) Assessing the resource allocation pattern in the health sector in AP.
   (c) Examine the feasibility of user fees as a source of financing in Andhra Pradesh district hospitals using a field study.

   In the next chapter, we take up the empirical analysis of evaluating hospital performance in the state of Andhra Pradesh using CUP analysis.