Health Economics

Dr. Pratap C. Mohanty

Department of Humanities and Social Sciences,

Indian Institute of Technology Roorkee

Week – 10

Lecture 48- Data Envelopment Analysis: Introduction

Welcome, friends. Once again, regarding our NPTEL MOOC module on Health Economics, We are in week number 10, which is unit number 10 on health efficiency. We have already started discussing the issues of productivity and efficiency. What are the concepts of this? Even in the previous lecture, we discussed various concepts of productivity and efficiency analysis in detail and their relevance. In this lecture, we will be sticking to Data Envelopment Analysis.

There will be couple of lectures on this, otherwise it is very difficult to give clarity on DEA that is Data Envelopment Analysis. It is highly applied irrespective of field. We have already discussed this in our previous lecture. Hence, our objectives in this class are measures of economic efficiency, input-oriented DEA, and output-oriented DEA, as well as a basic introduction to another model like SFA, which is stochastic frontier analysis.

So, what is that DEA, etc? We will start with some definitions and measures of economic efficiency. Then, we can address it correctly. While understanding economic efficiency, it refers to the optimal allocation of resources to maximize overall well-being or utility within an economy. So, it is related to optimal allocation to maximize the well-being of people.

This is used to assess how effectively resources are indeed applied to produce goods and services. There are two main components of economic efficiency, which are broadly called technical and allocative. We will also be evaluating and explaining to you how to measure technical efficiency and what the coefficient should be in our subsequent lecture through the application of the appropriate software. We will evaluate and emphasize the meaning of economic efficiency, which is jointly defined by technical and allocative efficiency. First of all, technical efficiency refers to the ability to avoid waste by producing according to the technology.

So, given the technology, how much can we address? On the other hand, the allocative component refers to the ability to combine inputs or outputs in optimal proportions while considering the prevailing prices. Given the price, hence it is referred to as the budget or the constraint. Given the constraint function in the production function, how can we best allocate or combine inputs to have optimal outputs that are all about the concept? So, the fundamental building block of the economic analysis of organizational efficiency is indeed the cost function or its counterpart, which is the production function.

So, either cost through optimization or production optimization are the two key important directions that we will be emphasizing. Koopman indeed in 1951 gave the formal definition, talks about producer is technically efficient if an increase in any output requires a reduction in at least one other output or increase at least one input. So, this means that technical efficiency, whenever we say, of course, we will emphasize increase in output, increase in output that is a function of reduction in at least other output if you have other targets as well if we have to reduce or we need to increase at least one input, then only we can achieve our target output. So, that is basically called technical efficiency. And if reduction in input required, this is in terms of output which was defined, if our target is through input minimization or the cost minimization, this requires to define as technical efficiency or technical efficient, this requires an increase in at least one or other input.

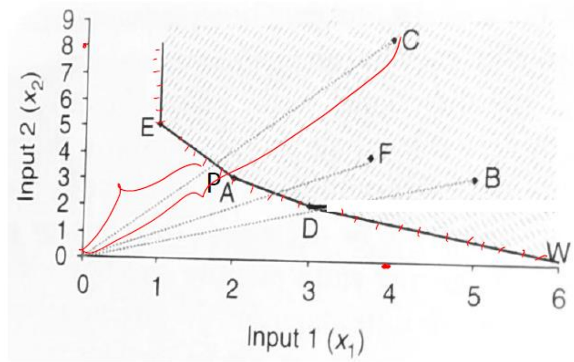
We can reduce the input requirements if and only if we need to increase the input or reduce at least one output. If you can reduce the output, then I think we can be in a position to optimize our input requirements. So, that is basically called technical efficiency. That is the formal definition from the original source by Koopman 1951. Debreu and Farrell also defined technical efficiency in other ways.

We have also cited in the previous lecture. They defined as input conserving orientation, largely called input oriented model and second one is called output augmenting orientation that is actually fostering output, this is also called output oriented efficiency, technical efficiency. Then another one is called non-oriented. We cannot even define if the already optimum level is reached and there is hardly any deviation; then, in that case, we may define the third one. Starting with input-conserving orientation, which is an input-oriented model.

This is defined as the maximum Equi proportionate reduction in all inputs that is feasible with a given technology and output, and this is usually identified in the radial distance. This is basically called the maximum Equi proportionate reduction in all inputs that is feasible with a given technology and output. This is also mentioned as keeping output fixed and exploring the proportional reduction in input usage, which is possible. How to construct a production frontier and the technical efficiency measured will be discussed shortly.

So, through this example (given below), here is an example our input are given as X_1 and X2 and output is Y and in this case there are 6 DMUs A, B, C, D, E, F and here the output is already considered to be neutralized as 1. Our combination of X_1, and X2 will lead to the maximum output possible, that is, if any, will define our technically efficient point. To find out the result, we will plot it accordingly in our diagram, but the first step is to understand through the input combination which you have taken proportionate to the target that is Y. So. Hence, we have taken X_1 by Y and X2 by Y, and we will plot them on the graph. Join the extreme point towards the center, which represents the minimum use of the input.

DMU	y	x_1	x_2
A	1	2	3
B	1	5	3
C	1	4	8
D	1	3	2
E	1	1	5
F	1	3.75	3.75



So, extreme points mean all sorts of combinations X_1, X_2, input 1, and input 2 we have taken by wherever the minimum possible basket is attached. So, we have plotted the line that is actually in a position to define the minimum most combination. However, this also derives the output target as 1. In this one, we will be emphasizing on the technical efficiency and that is simply defined as minimum input out of the actual input of the concerned DMU. So, in the specific DMU, we will define this technically efficient DMU by taking the minimum input, the less input out of the total or the actual input of the concerned DMU.

So, we will just see how it is defined. So, OP by OC is given; you can just have a look. Here is the diagram in the case of C, that is, the DMU C; you can see the input here is 4 and then $X_1 = 4$ and $X_2 = 8$; somewhere, we have plotted it a point of C. However, given the new radiant of C, we could easily trace that the optimum or the minimum input level is actually at P. Hence, we will define the OP as OP by OC.

$$TE_C = \text{Minimum input/Actual input of concerned DMU}$$

$$TE_C = OP/OC = d(OP)/d(OC)$$

* d= distance

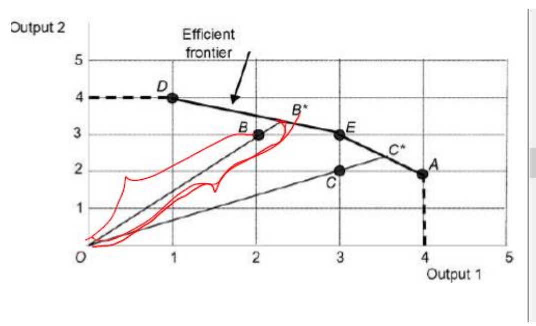
OP, the minimum distance is actually here by OC. This is the one. And so OP by OC, and here we are calculating the distance, D stands for the distance. So, the lowest is the case would be the better. Let us see if the firm or the DMU with the best combination and the firm or the DMU best in using input 1 can also be calculated.

Or the particular DMU which is actually using another input. So, wherever we can calculate the respective DMU with respect to 1 and with respect to 2. So, the best in using input 1 and the best in using input 2 can also be calculated. So, once again, we will clarify this through another diagram. At this stage, we have simply calculated the technical efficiency of the DMU, which is C.

In this diagram, DMU C we have calculated, and then we will compare which one is, in fact, the best. Now I will discuss something about output augmenting orientation and that is also called output-oriented efficiency model. This is defined as the maximum radial expansion in all output that is feasible. That means the maximum possibility will be achieved with the given technological inputs. So, this will keep an input constant and explore the proportional expansion in output quantity that is feasible.

How do we construct the production frontier and technical efficiency in this case? We need to assume, in that case, that we have to take the combination of Y_1 and Y_2 with a given level of input. Here, you can just mark the difference that the X is no longer considered to be constant. That can be variable as well. However, we can still find out the output augmenting technical efficient form or the DMU. So, let us see what is there in detail.

DMU	X	Y_1	Y_2
A	5	2	1
B	30	6	9
C	10	3	2
D	20	6	8
E	20	6	6



So, we need to follow the reverse. We used to define X1 divided by Y in the previous case and X_2 divided by Y. Since it is output-oriented, we now need to calculate Y_1 divided by X with respect to an input. Input is this with respect to X. As a ratio, we will actually find out, and then we can plot it accordingly. And then, since the outputs are more numerous and different, we can normalize outputs to 10 inputs in this example.

We can divide by 10 and accordingly calculate. In this example, that is whatever the ratio we got, and finally, we have divided that by 10 to normalize it. In this case, we will join the extreme point away from the center that represents the maximum possible production. Higher the boundary is higher the best output and that is basically called best practice frontier. It is also called an efficient frontier.

So, technical efficiency in the case of B DMU, we can just find out like this OB divided by OB*. So, the maximum possibility is OB*, and OB is their actual one. So, the technical efficiency of that case is OB divided by OB*. Then, the same distance formula is discussed, which we already discussed in the previous case. Third one is called non-oriented technical efficiency where no radial adjustment is possible.

It might be due to all DMUs already working efficiently. So, non-radial, non-oriented technical efficiency can be measured where no adjustment is actually required or possible. This kind of understanding can also be extended to the nth level if there are m inputs and m number of inputs, where we used to start with a limited number of inputs and 1 output or 2 output, 1 input. If it is m inputs and s outputs, then the production frontier becomes a surface in the dimension of m times s space.

So, repeating once again, we have derived that in this case, for the B-firm, B DMU is OB by OB*. So, mark the difference, and accordingly, we will also calculate and compare each of the DMUs and find out which one is indeed the best. So, what are the methods of measuring efficiency? How are people working? Usually there are these structures followed. For economic efficiency, parametric methods and non-parametric methods are used. Within parametric, there are frontier methods and non-frontier methods, and even in non-parametric methods also, the same two times are available.

In frontier methods, within the parametric structure, we have stochastic frontier analysis. In the case of non-frontier methods, we have a simple regression analysis where, since it is parametric, we will be deriving its parametric coefficients through the regression. In non-parametric method frontier analysis, the frontier methods part of the non-parametric analysis is called DEA, whereas you might have heard that DEA is a non-parametric method. As part of the frontier method, the borderline is mostly emphasized, and the optimum level is emphasized where DEA is discussed. The envelope of all the variables is discussed, and how far we are involved we are going to discuss in a short while.

In non-frontier methods, we will be using index numbers or in the case of a non-parametric structure. However, we are not emphasizing all sorts of things in this module; we are emphasizing the DEA as per our proposed direction. So, DEA we are emphasizing as I just said. Then what is parametric? I have already mentioned in the econometric approach to efficiency measurement, we usually assume the function and the production function might follow Cobb-Douglas one, maybe constant elasticity of substitution method or Translog function or maybe constant elasticity structure and transcendental function, etc.

There are different assumptions taken. For cross-sectional data, if you have two classes of econometric techniques available for efficiency analysis, one is called in case of cross-sectional data, one uses COLS that, is called corrected OLS, and stochastic frontier analysis SFA. We can also use this even in cross-sectional data. The difference between COLS and SFA lies with the interpretation of the residual and in COLS particularly the entire residual is interpreted as arising from inefficiency and in SFA the residual comprises a mixture of inefficiency and measurement error. In the case of non-parametric, as we already know, it does not have parameters, non-parametric piecewise convex isoquant is constructed such that no observed point lies to the left or below it and hence no functional form, and this is based on best practice frontier. Multiple inputs and outputs can also be considered, and this also provides information on Slack and Target.

These will be discussed in our subsequent lecture. An econometric approach to efficiency measurement we have to go by; if it is a parametric one, then we have already discussed it. So, accordingly, in the case of cross-sectional data, as I already mentioned, we have to take the help of COLS and SFA. Let us discuss DEA as part of the non-parametric efficiency analysis. So, data environment analysis is the most dominant approach for efficiency measurement, especially in healthcare and in many other sectors of the economy.

As mentioned by this author, DEA is a data-driven approach. The location of the efficiency frontier really matters while understanding the DEA. Highest ratio of output to input are considered efficient and the efficiency frontier is constructed by joining these observations up in the input output space. Inefficient organizations are usually enveloped by the efficiency frontier in DEA. So, whichever are actually inefficient organizations or the categories are actually enveloped together to derive results.

Hence, from the frontier, how far they are deviating can clearly be observed. DEA uses the DMU, which is a decision-making unit for each of the units of analysis; this term was coined by Charnes, Cooper, and Rhodes in 1978 in their seminal paper which introduced DEA. The DEA can reflect a whole range of different levels in healthcare settings. The entire healthcare system may be comparing countries, but there are different works we have cited for your readings. Health regions or health districts, hospitals, individual physicians, etc.

On this, DEA can be applied. We have also discussed this in our previous lecture. DEA is examined in various contexts, such as assuming constant returns to scale, production technology, variable returns, or even scale efficiency cases. So, in this lecture, we will first stick to the CES function, constant returns to scale, and CRS function. So, the DEA assesses efficiency in two stages. First, a frontier is indeed identified based on either those organizations using the lowest input mix to produce their outputs or those achieving the highest output mix given their inputs.

So, it is basically the lowest input mix to produce the target output as we usually do. Q is a function of X or where the target is either this maximize given this or minimize the C or maybe C you can write down C here, C minimizing C function of the level of Q. So, whoever direction we do, but our target is to actually achieve the lowest input mix or the highest output mix given their inputs. Hence, the models are largely conceptualized as input or output-oriented. Either our focus is to minimize input, or our focus is to maximize output.

Second, each organization is assigned an efficiency score by comparing its output-to-input ratio to that of an efficient organization that forms a piecewise linear envelope of surfaces in multi-dimensional space. So, basically we are linear enveloping those organizations which are away from the distance function or the radial distance and how far a score is defined and can be interpreted. Let us stick to input oriented efficiency. We will clarify one by one with the right example and its diagram. Suppose a DMU uses two inputs, X_1 and X_2, to produce a single output that is Y. In a healthcare setting, maybe a hospital uses two inputs, such as doctors and nurses, to produce a single output, that is, a patient's treatment.

Assuming diminishing marginal factors of productivity, isoquants can be constructed so that those will be convex to the origin. Along the frontier, reduce the use of one input, say X1, that is, doctor or maybe another input, that is, nurses, unnecessary increase in the use of other input, that is, extra nurses, to maintain the level of treatment provided. So, if we are reducing one, we may have to increase another one in order to achieve the target output, which is Y, which is treatment. So, this is what is given. Hence a curve that is we are deriving a production possibility through isoquant since we are plotting it through the inputs, two inputs.

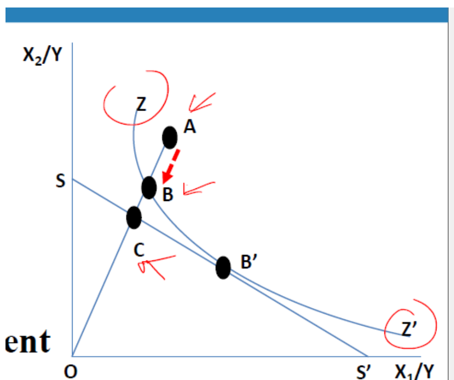


Fig.- Technical and allocative efficiency under an input orientation

Hence, our maximum possible bundles are presented through ZZ'. So, here it is, this is ZZ', and this is convex to the origin since we are diminishing marginal product is followed. All DMUs lie on the production frontier if they are actually efficient or above it if they are inefficient. If even higher cost is actually required to produce, then, of course, that is not considered to be efficient. So, we will be emphasizing technical as well as allocative efficiency one by one.

At this time, we are discussing how Hospital A could proportionately reduce its use of doctors and nurses, given the amount of treatment it provides, and move to a feasible and technically efficient production point such as that adopted by Hospital B. So, there are DMU units basically A, B, C are presented in this case A, B, C. You can just have a look here are the units. We will be comparing and explaining the possibilities. SS prime is a budget line or the isocost line, which reflects the ratio prices of the inputs and. That when we are taking ratio of the prices we are supposed to take up the discussion of allocative efficiency since the prices of inputs are derived then the best use of inputs or the allocation of them is a matter of concern for the industry or for the organizations hence allocative efficiency will be discussed.

At this stage, we are sticking to the output given input. We are trying to reach the boundary line, maybe through a combination of input to output or a combination of output to the minimum level of input we will be emphasizing in a short while. The cost-efficient point of production, in this case, is actually B prime since that is tangential, and we have already read in our microeconomic theory that when it is tangential, the budget line is tangential to that of the isoquant, and or the cost line is tangential, to the cost line S, S prime is cost line we reach the optimum level of the production or the cost-effective point of production. Thus, in that case, hospital S production is currently much above the production frontier, which is inefficient, and that is, more inputs are required to produce the output at point A. It is actually inefficient. It is much above, and in the radiant, you can just check that it is much above to achieve, and so more inputs are required to produce the output at A than if it were to move to a point such as B or B prime. So, let us see in DEA technical inefficiency is usually made using the radial measure of inefficiency by comparing where the hospital is in relation to the production function that is the distance of B to A. This distance is actually unnecessarily over expensive or over inputs are required to get the treatment.

So, the treatment actually lies with the Z and Z' production possibility or the isoquant where the output target is actually if you remember our example X_1, X_2 and Y. So, Y is actually required B to A is inefficiently I look at I mean used that has to be controlled. So, the distance BA is the amount by which all inputs of doctors and nurses could be proportionally reduced without a reduction in hospital admission. So, this one could be proportionally reduced to this level; hence, we can reach it at an optimal point.

This is expressed in percentage terms by the BA by OA. This is basically an inefficient proportion. This proportion, out of the total combinations of input and output, this distance could be actually minimized, and hence, it is actually called. So, then, how do we measure technical efficiency in this case, which is the oriented model we have been emphasizing? So, in this case, for hospital A, we have to define that OB out of OB by OA. OB is actually the optimum one so far as the isoquant is concerned, and OA is its own allocation of factors.

Hence it is the technical efficiency through input oriented model of hospital A is defined as OB upon OA. This is also equal to 1 minus inefficient technical appreciation, which is BA by OA, and 1 minus BA by BA is the inefficient allocation. So, if that extent is minimized will be actually useful. Then what do you mean by pure technical efficiency? This shows the deviation from the production frontier that is Z', and this value lies between 0 and 1 if 1 by any deviation if we are achieving its value, lies between 0 and 1; if it is 1, that means full technical efficiency is achieved if hospital A produced at a point such as B. Once the B level is achieved that means full technical efficiency is achieved.

Then, if the isocost can be specified because the input prices are known, then that way, when isocost is defined very clearly here, it is SS'. If that is very clearly defined and then the respective prices of X_1 and X_2 are clearly specified, in that case, we are actually referring to allocative efficiency, and in short, it is called allocative efficiency through the input-oriented model. So far, we have been discussing technical efficiency; you can just have a mark; we have been discussing technical efficiency. When we are responding through the best possible choice, which is through the inputs and their cost functions, then we will be emphasizing their allocative efficiency. In this case of the operator, if the hospital we are referring to the hospital A or the DMU unit A, referring to point A as follows.

So, allocative efficiency through input output oriented model is OC upon OB. So, if cost is known for this particular case, the ratio would actually be OC by OB. The optimum one of course will be at B, hence the allocative one has to be OC. Where the distance CB, CB is the reduction in production cost that would actually be required..

This reduction is considered to be good. The reduction in production cost that would occur if production were to take place allocatively and technically. So, the allocative efficient point is B prime. So, since at B', we have already seen that at B prime and at C, the cost levels are the same, they are on the same isocost line, and the maximum production is also achieved. Hence, the distance CB is the reduction in the production cost that would occur if production were to take place at the allocatively efficient point that is B prime instead of the technically efficient point, but allocatively efficient point that is point B. So, I will further clarify that it does represent the deviation from the price efficient point.

So, we are now defining the price efficient point as that of the production efficient point. The extent to which DMU incurs expenditures that are more than the minimum feasible one, therefore, comprises two components. One is called technical efficiency, and the other is called allocative efficiency. We have already clarified. The production of these measures can be combined together that is basically called total economic efficiency (EE input oriented) that is economic efficiency, total economic efficiency input oriented model that will be simply multiplication of technical efficiency input oriented and allocative efficiency input oriented.

So, in the case of this technical efficiency, we have already defined OB divided by OA. Here is the point OB divided by OA, and then, in the case of allocative, one OC is divided by OB. So, finally, this is derived as OC divided by OA. So, the economically efficient allocation would be an efficiency combination that will be OC upon OA. So, similarly for all DMUs we can able to easily calculate their possibilities and then we can compare which one is really efficient.

$$EE_{IN} = TE_{IN} \times AE_{IN}$$
$$= OB/OA \times OC/OB$$
$$= OC/OA$$

Another is called output-oriented. So far, we have been emphasizing what is called an oriented model. When the focus is output-oriented, of course, the target is actually output-oriented. So, in that case, if there are two outputs, then how best is the output to be utilized so that it will be efficient? In this case example, you have taken inpatient treatments and outpatient visits.

There are two outputs for a single input that is hospital staff. Given the hospital staff, how best can we allocate inpatient treatment and output outpatient visits so that our model is considered to be efficient? So, once again, here it is. ZZ' is called the production possibility frontier. Why is it called the production possibility frontier? We have taken this into account with respect to the output. In the earlier case, it was with respect to inputs.

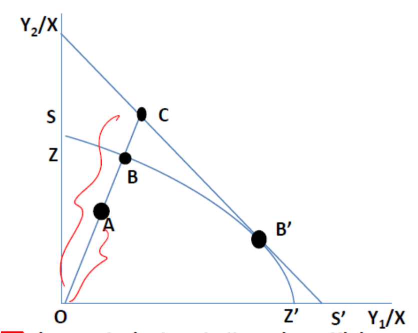


Fig.- Technical and allocative efficiency under an output orientation

So, isoquant was derived, and now it is a production possibility frontier. So, all hospitals lie on the production frontier if they are efficient or below it. If they are below, then, of course, they are inefficient. Using the output orientation, hospitals that lie below the production frontier, that is A, in this case, hospital A could proportionally expand their output quantities or basically Y_1 and Y_2 of inpatient treatment and outpatient visits. The holding level of input used is X, or that is hospital staff constant.

Given the constant level of this, they can increase proportionately. Under the existing technology, they could do this up to a point such as B, which might rise to this level, which is located on the production boundary. If information about the relative value of two outputs is available, then we can also discuss the possible constructs to construct the equivalent of an iso-revenue line. So, that is also possible. So, SS' is our line iso-revenue curve that reflects the market value of the two outputs, Y_1 and Y_2, usually called the price line or the contract curve. The efficient point of production is, in fact, B prime, which is highlighted over here on the tangential point between iso-revenue and the production possibility frontier and the technical efficiency in this case in the output-oriented model.

Hence, we have written that, as technical efficiency output oriented of hospital A, it is, of course, OA by OB as simple as that OA by OB. Whereas the allocative efficiency, which is done now, is just the reverse, we are actually sticking to the output-oriented model, and the ratio would be accordingly taken. So, that is, in fact, OB for the technical efficiency one OB by OC or focus is actually on this the iso-revenue line, and that revenue has to be attended. So, OB upon OC is the allocative efficiency output-oriented model. Hence, the total economic efficiency can also be calculated by multiplying these two, and finally, it results as OA upon OC.

$$
\begin{aligned}
\mathbf{EE_{OUT}} &= \mathbf{TE_{OUT} \times AE_{OUT}} \\
&= \mathbf{OA/OB \times OB/OC} \\
&= \mathbf{OA/OC}
\end{aligned}
$$

So, this is indeed called the total economic efficiency. In healthcare, output prices are seldom available, so most studies restrict the analysis to the calculation of technical efficiency. Since the output prices are hardly available and hence not the total economic efficiency, it is very difficult to calculate in healthcare research. So, technical efficiencies is mostly restricted in research. All these measures of efficiency, that is, technical, allocative, and economic efficiency, are bounded by 0 and 1.

We will discuss this value in our calculation using the right software. We will be using DEAP software as well. And that is open-access software. These radial efficiency measures are units invariant. So, this is scaled down in the variables, and hence, those values are considered to be comparable. So, these are all different calculations of technical efficiency and allocative efficiency both in the oriented model and the oriented model concerned.

I think you will get the applied version in our subsequent lecture. And these are our further readings. I suggest you read this for further clarification. And in the next lecture, we will be emphasizing variable returns to scale model and constant returns to scale. We will also compare constant returns to the scale model versus variable returns to the scale model, which is short and called the CCR and BCC models in the DEA framework and scale efficiency. We will also deal with some slacks and how slacks are defined. With this, I will stop here. I look forward to your participation. Thank you.