Data Envelopment Analysis (DEA):
Introduction, Definitions, and Examples

Let us label two clinics within the same healthcare system A and B. Both clinics have common information systems, which include common modules for accounting, scheduling, and billing. In addition, both clinics focus on outpatient services for patients with Diabetes, handle the same number of patient visits per year, and have identical staff levels.

A system manager might view profits and waiting times as an outcome of how the system is managed. Clinic A is profitable and Clinic B operates at a substantial annual loss. In addition, Clinic A has much shorter waiting times than Clinic B. The manager asks whether Clinic A being managed better than Clinic B.

The obvious answer to this question is "it depends." A closer look at the patient populations reveals that Medicare is the payer for 10% of the patients seen in Clinic A, but is the payer for 50% of the patients seen in Clinic B.

Some patients with diabetes eventually develop kidney failure. When this happens within this system any patients identified with this co-morbidity in Clinic A are sent to Clinic B for all future visits and these visits take longer to complete. Thus we may say that Clinic A has different inputs to consider including the payer, and case mix. Thus, we conclude that metrics of productivity or efficiency that ignore these factors are of limited use to the system manager.

The chief objective of this note[1] is to introduce the basic elements of Data Envelopment Analysis as it is applied to benchmarking units that perform similar functions in settings that vary within the Healthcare industry.

Introduction:
Consider an engine in an automobile. One common way to discuss the efficiency of that engine is to speak in terms of the Miles Per Gallon (MPG). Each potential customer in the US is quite familiar with this measurement and the units involved. We can understand the notion that traveling more miles using a single gallon of gas is better – *if all other things are equal*.

This last phrase instantly brings a host of complications to the discussion. Simply put, all other things are NOT equal. Some vehicles are relatively large. Others are relatively fast. Some are designed to carry large loads, while others are designed to be luxurious. The basic question that we wish to deal with in this note is, "How can we discuss efficiency in a coherent manner when the components of the definition have multiple dimensions?"

This question is particularly prevalent in the healthcare sector. For example, we may need to consider the efficiency of a clinic, a hospital or an insurance provider. In many of these cases, simply looking at profit and loss statements misses many critical aspects of the importance and complexity of the

[1] May 2018. Chester Chambers prepared this note as a basis for classroom discussion. Please do not duplicate without authorization.

service or activity. For Clinics A and B, the fact that Clinic A is more profitable than Clinic B, does not necessarily tell us that one clinic is being managed better than the other. Differences in financial results may easily stem from differences in payers or case mix. They may also differ in their teaching mission, regulatory constraints, or fixed costs such as rent. The intuitively simple concept of efficiency quickly becomes quite complex when real people in real locations, dealing with real patients are involved.

DEA is a tool, which has proven useful in that it provides a way to deal with three particularly difficult questions in discussions of efficiency. To come up with a measure of efficiency we need to understand something about the inputs to, and outputs from an operating system. First, what are the relevant inputs to the system? These may include direct labor hours, the dollar value of materials used, the number of supervisors, the average level of experience among the staff, the physical space occupied, etc. It may also include the severity of the conditions treated, demographic characteristics of the patient pools, and even payers. This begs the question of, "how are we going to measure these inputs when they are not easily convertible to dollar figures?".

Second, what are the appropriate outputs of the system? These may include jobs processed, or customers served. However, it may also include patient satisfaction, average waiting time, complications after discharge, etc. Again we see that monetary units are not always available, and sometimes are completely inappropriate.

Third, what are the appropriate ways of measuring the relationship between these inputs and outputs?

Measuring Service Productivity:
The measure of an organization's productivity, if viewed from an engineering perspective, is similar to the measure of a system's efficiency. It can be stated as a ratio of outputs to inputs (e.g. miles per gallon, cases per hour, discharges per day, etc.)

To evaluate the operational efficiency of a clinic, a simple ratio such as revenue/cost might be used. However, too much focus on an accounting measurement that ignores complicating factors may motivate behavior, which leads to problems in the future. For example, working to reduce costs by reducing face time with a physician may result in patient behavior that increases the chances of complications tomorrow, and this is certainly not a good idea.

Stated more generally, a major problem with using simple ratios is that the results may be uninformative if the mix of outputs is not considered explicitly. For example, an understanding of readmissions or complications is essential to assess clinic performance from both a medical and economic perspective. This same criticism also can be made concerning the mix of inputs.

The DEA Model:
Fortunately, a technique has been developed with the ability to compare the efficiency of multiple operating units that provide similar services by explicitly considering their use of multiple inputs (i.e. resources, patients, payers) to produce multiple outputs (i.e. clinic visits, patient satisfaction, quality of life). Data Envelopment Analysis (DEA) circumvents the need to develop standard costs for each service because it can incorporate multiple inputs and outputs in both the numerator and denominator of an efficiency ratio without the need for

conversion to a common dollar basis. Thus, the DEA measure of efficiency explicitly accounts for the mix of inputs and outputs and, consequently, is more comprehensive and reliable than a set of operating ratios or profit measures.

DEA makes use of a Linear Programing (LP) model to draw comparisons among a set of service units. (In two dimensions a Linear model involves straight lines and we often will use this feature to display our results.) The model attempts to maximize an operating unit's efficiency score, expressed as a ratio of outputs to inputs, by comparing a particular Decision Making Unit's (DMU's) efficiency score with that of a group of similar DMUs that are delivering the same types of service. In the process, some DMUs will be described as achieving 100% efficiency. These are referred to as the *relatively efficient units*, whereas other units with efficiency ratings of less than 100% are referred to as *inefficient units*. The term DMU is being used only because this term was used in the original research on DEA.

Technically, speaking the phrase 100% efficiency may be a bit misleading. Being perfectly efficient is physically impossible. What we are identifying in actuality is the set of service units about which it cannot be said that other units would result in a higher efficiency rating if the inputs and outputs are weighted consistently across all of the service units in a given set.

Managers can use DEA to compare a group of service units to identify relatively inefficient units, measure the magnitude of the inefficiencies, and by comparing the inefficient with the efficient ones, discover ways to reduce those inefficiencies.

The DEA linear programming model that we will use here is quite simple but powerful. It has been modified in several ways to handle problem complexities that we will not discuss here. But the basic model that we present is the starting point of most DEA approaches.

Definition of Variables:
Let E_k, with $k = 1,2,....,K$ be the efficiency ratio of unit k, where K is the total number of units being evaluated.

Let u_j, with $j=1,2,...,M$, be a coefficient of output j, where M is the total number of output types considered. The variable u_j is a measure of the relative decrease in efficiency with each unit reduction of output value. For example if a clinic is being considered and the 2nd of the outputs being measured is patient satisfaction scores, then u_2 is equivalent to the drop in the efficiency rating for this clinic if the average patient satisfaction score drops by 1 unit.

Let v_i, with $i=1,2,...,N$, be a coefficient for input i, where N is the total number of input types considered. The variable v_i is a measure of the relative increase in efficiency with each unit reduction of input value. For example, if the third input measured was nursing-hours then v_3 measures the increase in efficiency if the same output could be produced with 1 fewer hour of nursing labor.

Let O_{jk} be the number of observed units of output j generated by service unit k during one time period. For example, patients treated, prescriptions filled, or insurance policies audited.

Let I_{ik} be the number of actual units of input i used by service unit k during one time period. Examples may include, nursing hours used, examination rooms used, or years of experience among the staff.

Objective Function:
The objective is to find the set of coefficients associated with each type of output (u_j values) and coefficients associated with each type of input (v_i values) that will give the service unit being evaluated the highest possible efficiency.

$$\max E_k = \frac{u_1 O_{1k} + u_2 O_{2k} + \ldots u_M O_{Mk}}{v_1 I_{1k} + v_2 I_{2k} + \ldots v_N I_{Nk}} \qquad (1)$$

where k is the index of the unit being evaluated. Note that the input levels and output levels (I and O values) will be observed directly. What we need to find are the weights that we will assign to each input and output (u, and v values) when calculating the "score" for each unit.

Constraints:
This objective function is subject to the constraint that when the same set of input and output coefficients $\left(u_j\text{'s and } v_i\text{'s}\right)$ is applied to all other service units being compared, no service unit will receive an efficiency rating which exceeds 100% or a ratio of 1.0.

$$\frac{u_1 O_{1k} + u_2 O_{2k} + \ldots u_M O_{Mk}}{v_1 I_{1k} + v_2 I_{2k} + \ldots v_N I_{Nk}} \le 1.0, \qquad (2)$$

Where $k = 1,2,\ldots,K$ and all coefficient values are positive and nonzero.

This type of problem is surprisingly difficult to solve when stated in this form. The reason is that it is non-linear. The fact that decision variables are divided by each other also makes it difficult to show how the method works. Therefore, we will make a slight reformulation by rearranging terms a bit.

Note that both the objective function and all constraints are ratios rather than simple linear equations. The objective function in equation (1) is restated as a linear function by scaling the weighted inputs for the unit under evaluation to a sum of 1.0.

Speaking more intuitively, when we focus on any particular unit we can fix its inputs at their current level and scale that value to equal 100% or simply 1. This makes the denominator of equation (2) equal to 1, which means that we can restate the objective as,

$$\max E_e = u_1 O_{1e} + u_2 O_{2e} + \ldots u_M O_{Me} \qquad (3)$$

subject to the constraint that,

$$v_1 I_{1e} + v_2 I_{2e} + \ldots v_N I_{Ne} = 1 \qquad (4)$$

This turns out to be a much easier problem to present and to solve. For each service unit, the constraint in equation (2) is similarly reformulated as:

$$\begin{aligned} & u_1 O_{1k} + u_2 O_{2k} + \ldots u_M O_{Mk} \\ & -\left(v_1 I_{1k} + v_2 I_{2k} + \ldots v_N I_{Nk}\right) \le 0 \end{aligned} \qquad (5)$$

Where $k = 1,2,\ldots,K$ and,

$$\begin{aligned} u_j &\ge 0 \quad j = 1,2\ldots,M \\ v_i &\ge 0 \quad i = 1,2,\ldots,N \end{aligned}$$

In other words, if A/B < 1, we can rewrite this as A < B or as A – B < 0. Thus a complex looking ratio becomes a simple difference.

A question of sample size often is raised concerning the number of service units that are required compared with the number of input and output variables selected in the analysis. The following relationship among the number of service units K used in the analysis and the number of input N and output M types being considered is based on empirical findings and the experience of DEA practitioners.

$$K \geq 2(N+M) \qquad (6)$$

For example, if we consider a setting with 2 types of inputs and 1 type of output, we want to have at least 2(2 + 1) = 6 units as a set of sites to work with.

Example: Med-Lab Inc.[2]
An innovative chain has six units in several different cities, which provide testing services on an on-demand basis. Only a standard blood test is available. (This assumption is not necessary for DEA. It is made here only to make it easier to discuss a graphical representation of model results.) Management has decided to use DEA to improve productivity by identifying which units are using their resources most efficiently and then sharing their experience and knowledge with the less efficient locations. Table 1 summarizes the data considering two inputs: labor-hours and material dollars consumed during a typical period to generate an output of 100 completed tests.

Table 1:

Service Unit	Tests Done	Labor Hours	Material Dollars
1	100	2	200
2	100	4	150
3	100	4	100
4	100	6	100
5	100	8	80
6	100	10	0

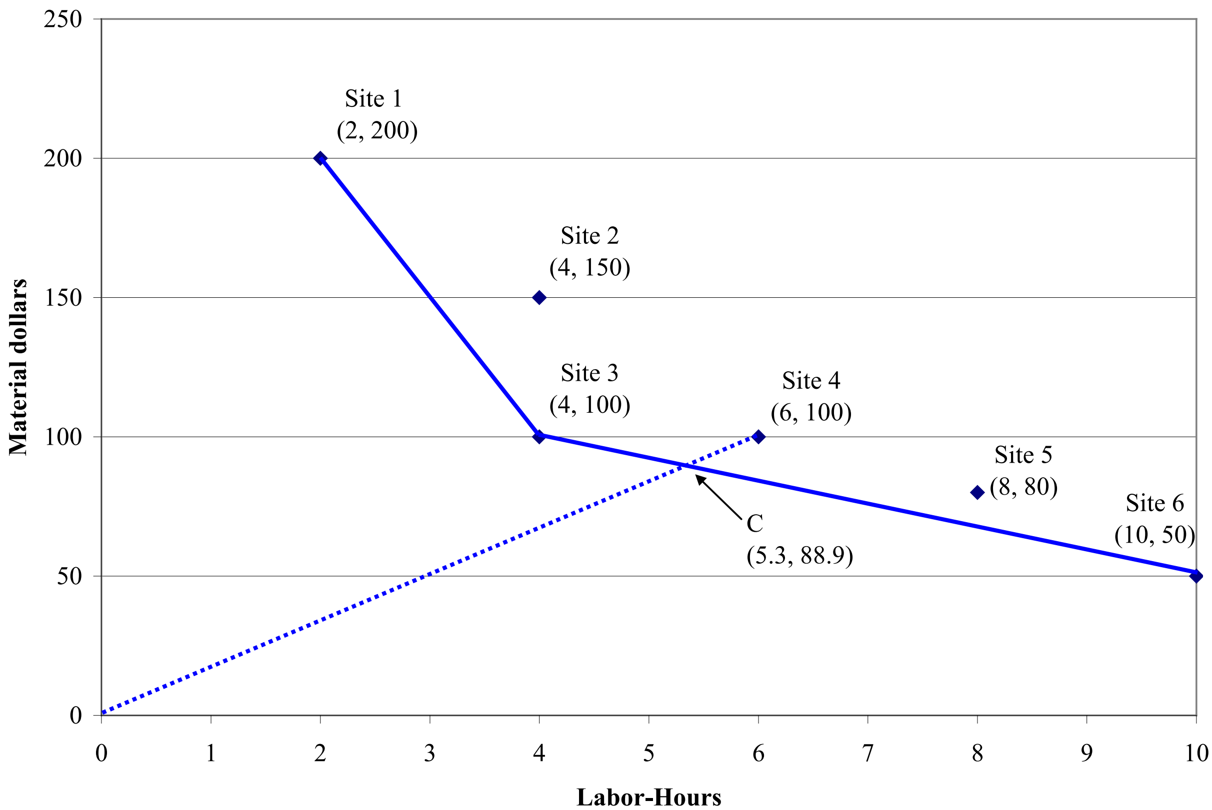
As Figure 1 shows, service units S_1, S_3, *and* S_6 have been joined to form an efficient production frontier of alternative methods of using labor hours and material resources to complete 100 tests. As can be seen, these efficient units define an "envelope" that contains all the inefficient units – thus the name "Data Envelopment Analysis."

Note that in most cases output will vary among the service units. We are dealing with these differences in scale by focusing on the inputs needed per 100 tests done. For example if unit 1 actually only completed 50 tests with 1 labor hour and 100 dollars in material, we scale this unit's output to make it comparable to the other units. Thus we discuss all units on the same basis by considering inputs per 100 tests done. Again, this is not necessary for DEA to work. It is being done here only to make the resulting model very simple to interpret and easy to display.

For this simple example, we can identify efficient units by inspection and see the excess inputs being used by inefficient units (e.g. S_2 would be as efficient as S_3 if it used \$50 less in materials.) To gain an understanding of DEA, however, we will proceed to formulate the linear programming problems for each unit, then solve each of them to determine efficiency ratings and other information.

[2] This example is adapted from *Service Management* by J.A. and M.J. Fitzsimmons, ©2008.

Figure 1: Productivity Frontier of Med-Lab



We begin by illustrating the LP Formulation for the first service unit, S_1, using equations (3), (4), and (5).

$$\max E(S_1) = u_1 100$$

subject to

$$u_1 100 - v_1 2 - v_2 200 \le 0$$
$$u_1 100 - v_1 4 - v_2 150 \le 0$$
$$u_1 100 - v_1 4 - v_2 100 \le 0$$
$$u_1 100 - v_1 6 - v_2 100 \le 0$$
$$u_1 100 - v_1 8 - v_2 80 \le 0$$
$$u_1 100 - v_1 10 - v_2 50 \le 0$$
$$v_1 2 + v_2 200 = 1$$
$$u_1, v_1, v_2 \ge 0$$

Let us discuss this LP in more explicit terms. We want to determine coefficients that maximize the efficiency of this unit. This efficiency is defined by a total measure of outputs (100 tests) divided by a total measure of inputs (2 labor hours for site 1 and 200 dollars in materials.) When focused on a single location, we can scale the inputs to equal 1. Therefore, the objective simply becomes to maximize the weighted output, given 100% of the site's inputs. Formally, this is simply (u_1 * 100)/1.

Next, since it is physically impossible to have an efficiency rating above 100% = 1, we have,

$$\frac{u_1 100}{v_1 2 + v_2 200} \leq 1 \; or,$$
$$u_1 100 \leq v_1 2 + v_2 200 \; or,$$
$$u_1 100 - v_1 2 - v_2 200 \leq 0.$$

In other words if weights u_1, v_1, and v_2 are used efficiency cannot be above 100% for unit 1. A parallel statement is made for each of the other locations. In other words, if weights u_1, v_1, and v_2 are used for any unit efficiency cannot be above 100% for that unit either. This explains the first 6 constraints of the model.

The last constraint simply reflects the fact that we are scaling the inputs at the site under consideration to equal 1 as explained earlier.

To gain some insight into the practical value of DEA, let us consider the problem from a managerial perspective. If I have oversight of multiple units it is clear that I want the efficiency of each unit to be 100%, or as close as possible. However, since there are multiple inputs to the process it becomes difficult to compare units. The manager of a unit that uses very little labor but a lot of material will argue that this site is making an efficient use of labor and should be evaluated accordingly. Another manager of a unit that uses more labor but less material will argue that that unit is making an efficient use of material and should be given credit for this.

In effect, what this DEA model does is it identifies weights that we will apply to the levels of resources used by the manager of Site 1 that results in the greatest possible efficiency calculation for that site. The one caveat is that the weights that we use for that site cannot result in an efficiency calculation which exceeds 100% for any other site since this is physically impossible. In effect, we will look at Site 1 in the most favorable light possible that does not violate common sense. The manager of a site deemed to be less than 100% efficient using this method is not likely to be thrilled by this result. But it is impossible for him/her to say that any other set of weights would make the site look any better. Consequently, some other explanation for the outcome must be identified.

Similar linear programming problems are formulated (or better yet, the S_1 linear programming problem is edited) and solved for each of the other service units by substituting the appropriate output function for that site and substituting the appropriate input function for the last constraint. Constraints 1 through 6 which restrict all units to no more than 100% efficiency remain the same in all cases.

The set of six linear programming problems was solved with Excel Solver. (The Excel spreadsheet is provided to accompany this note.) Notice that when considering each unit, only the last constraint must be edited by substituting the appropriate labor and material input values from Table 1 for the unit being evaluated. The solution to the LP for unit 1 is shown in Table 2.

Table 2:

Unit	1					
	U1	V1	V2	Eff.		
Value	0.010	0.167	0.003	1		
Unit 1	100	-2	-200	=	0.000<=	0
Unit 2	100	-4	-150	=	-0.167<=	0
Unit 3	100	-4	-100	=	0.000<=	0
Unit 4	100	-6	-100	=	-0.333<=	0
Unit 5	100	-8	-80	=	-0.600<=	0
Unit 6	100	-10	-50	=	-0.833<=	0
Inputs	0	2	200	=	1.000=	1

Note that the decision variables are u_1, v_1, and v_2. Also recall that each site produces 100 tests. Therefore the solution shown states that the efficiency of Unit 1 is u_1 * (100) = 0.01 * (100) = 1, or 100%. The values for v_1 and v_2 that are associated with the inputs of labor hours and materials respectively measure the relative decrease in efficiency with each unit increase of an input. Thus we see that if Site 1 produced the same output using one additional hour of labor, its efficiency would be 0.167 = 16.7% lower. Similarly, if it produced the same output consuming one dollar more in materials, it would decrease its efficiency by 0.003 = 0.3%.

Table 3 shows the result of a parallel analysis for unit 4.

Table 3:

Unit	4					
	U1	V1	V2	Eff.		
Value	0.009	0.056	0.007	**89%**		
Unit 1	100	-2	-200	=	-0.556<=	0
Unit 2	100	-4	-150	=	-0.333<=	0
Unit 3	100	-4	-100	=	0.000<=	0
Unit 4	100	-6	-100	=	-0.111<=	0
Unit 5	100	-8	-80	=	-0.089<=	0
Unit 6	100	-10	-50	=	0.000<=	0
Inputs	0	6	100	=	1.000=	1

We immediately notice that the efficiency of this site is only 89%. This motivates the manager to identify actions or policies to increase the efficiency of unit 4. By considering the values of v_1 and v_2 we see that each unit decrease in labor hours results in an efficiency increase of 0.056. In order for unit S_4 to become relatively efficient it would have to increase its efficiency rating by 0.11 points or 11 % (100% - 89%). This could be accomplished by reducing the labor hours used by roughly 2. We know this is true because 2 * v_1 = 2 * 0.056 = 0.112 or about 11%. Note that with this reduction in labor hours, unit S_4 becomes identical to the efficient unit S_3.

The coefficient related to the dollar value of materials used (v_2) is 0.0067. This means that an alternative approach to make S_4 100% efficient would be to produce the same output with a reduction in materials used of 0.111/0.0067 = $16.57.

It is very important to recognize that these are only two of an infinite number of approaches to achieve 100% efficiency for this unit. Any linear combination of changes to these two resource levels that produces the 0.11 increase in the efficiency rating would also make unit S_4 efficient.

Considering the graphic shown in Figure 1, we see that one path to efficiency is to move from the current point S_4 along the dotted line until it intersects with the efficient frontier defined by the line segment joining efficient units S_3 and S_6. In some sense, this is also the shortest path to efficiency for this unit.

Table 4 shows the values of u_1, v_1, and v_2 for all six units.

Table 4:

Unit	u1	v1	v2	Eff.
1	0.01000	0.17	0.003	100.0%
2	0.00857	0.14	0.003	85.7%
3	0.01000	0.06	0.008	100.0%
4	0.00889	0.06	0.007	88.9%
5	0.00909	0.06	0.007	90.9%
6	0.01000	0.06	0.008	100.0%

In Table 4 we find that DEA has identified the same units shown as being efficient in Figure 1. Units S_2, S_4, and S_5 all are inefficient in varying degrees.

Table 5 shows a report that Excel can generate any time it solves a Linear Programming problem such as we have here. Excel labels this the Sensitivity report.

Table 5: Sensitivity Report from LP for unit S_4

Microsoft Excel 11.0 Sensitivity Report
Worksheet: [DEAExample.xls]Sheet1
Report Created: 3/3/2008 6:01:14 PM

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
$B$11	Value U1	0.008888889	0	100	1E+30	100
$C$11	Value V1	0.055555556	0	0	2	7
$D$11	Value V2	0.006666667	0	0	116.6666667	33.33333333

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
$F$14	=	-0.556	0.000	0	1E+30	0.555555556
$F$15	=	-0.333	0.000	0	1E+30	0.333333333
$F$16	=	0.000	0.778	0	0.142857143	0.5
$F$17	=	-0.111	0.000	0	1E+30	0.111111111
$F$18	=	-0.089	0.000	0	1E+30	0.088888889
$F$19	=	0.000	0.222	0	0.153846154	0.625
$F$20	=	1.000	0.889	1	1E+30	1

Let's translate this report into more simple English. First "Adjustable Cells" refer to the values that we asked Excel to find. These are the weights applied to the inputs and outputs listed. The "Final Value(s)" are the numerical values that Excel found which maximize the objective function while satisfying all constraints.

We are particularly interested in the column titled, "Shadow Price". A shadow price is linked to each constraint and it is either 0 or some positive value. Having a shadow price of 0 means that relaxing this constraint will not change the optimal solution. The only that this can be true is if this constraint is not part of the set of constraints defining the optimal solution. In other words, if the shadow price is 0, then this constraint is not binding at the optimal solution. If the shadow price is positive, then we know that relaxing this constraint allows for an increase in the objective value. This is only possible if this constraint is in the set of constraints defining the optimal solution. In other words this constraint is binding.

The shadow price values of 0.778 and 0.222 correspond to the constraints associated with units S_3 and S_6. These non-zero shadow prices tell us that when maximizing the efficiency of unit S_4 the LP could not reach a level of 100% without using weights that would suggest that units S_3 and S_6 have efficiency levels above 100%.

Using the "language" of DEA we would say that the inefficient unit S_4 should be evaluated through comparison to this "efficiency reference set." Visually we illustrate this by noting that the efficient units S_3 and S_6 have been joined with a line defining an "efficient frontier". A dashed line drawn from the origin to inefficient unit S_4 cuts through this frontier and thus shows that unit S_4 is inefficient.

This use of shadow prices to generate a relevant reference set for each inefficient unit is an attractive bi-product of the DEA approach. When only 2 dimensions are involved, this may seem like overkill. It is easy to see which efficient units are "close to" the inefficient unit, and common sense may suggest that we look at what is being done at these locations for clues on how to improve the relatively poor performer. However, for more complex settings with multiple inputs and multiple outputs, the identification of proper points of comparison for inefficient units can be extremely helpful.

Table 6 contains calculations for a hypothetical unit C, which is a composite reference unit defined by the weighted inputs of the reference set S_3 and S_6. As Figure 1 shows, this composite unit C is located at the intersection of the productivity frontier and the dashed line drawn from the origin to unit S_4. Thus, compared with this reference unit C, inefficient unit S_4 is using excess inputs in the amounts of 0.7 labor hours (6 – 5.3) and 11.1 material dollars (100 – 88.9). Intuitively, this says that the most direct path to 100% efficiency for unit S_4 is to cut labor hours by 0.7 and material costs by 11.1.

Multiple Inputs and Outputs:

One can argue that DEA was not really needed for the previous example. In very simple cases DEA adds complexity to the analysis of a setting, which is rather straight forward and nobody likes a showoff. If there is only one output, any act that increases the output that costs less than the value of that increase must be valuable. Let us turn to a slightly more complex and more realistic example.

Mike Lister is a district manager for the Med-Labs[3] chain. The region Mike manages contains 12 company-owned units. Mike is in the process of evaluating the performance of these units during the past year in order to make recommendations on how much of an annual bonus to pay each unit's manager. He would like to base this decision, in part, on how efficiently each unit has been operated. Mike has collected the data shown below on each of the 12 units. The outputs he has chosen include each unit's net profit (in $100,000's), average customer satisfaction rating, and average monthly cleanliness score. The inputs include total labor hours (in 100,000's) and total operating costs (in $1,000,000s). He would like to apply DEA to this data to determine an efficiency score for each unit.

[3] This example is adapted from *Spreadsheet Modeling and Decision Analysis* (2001) by C.T. Ragsdale.

Table 6: Calculation of Excess Inputs Used by Unit S_4

Outputs and Inputs	Reference Set	Component Reference Unit C	S_4	Excess Inputs Used
Meals	**(0.778) * 100 + (0.222) * 100 =**	**100**	**100**	**0**
Labor-hours	**(0.778) * 4 + (0.222) * 10 =**	**5.3**	**6**	**0.7**
Material ($)	**(0.778) * 100 + (0.222) * 50 =**	**88.9**	**100**	**11.1**

Table 7:

	Outputs			Inputs	
Unit	**Profit**	**Satisfaction**	**Cleanliness**	**Labor-Hours**	**Operating Costs**
1	5.98	7.7	92	4.74	6.75
2	7.18	9.7	99	6.38	7.42
3	4.97	9.3	98	5.04	6.35
4	5.32	7.7	87	3.61	6.34
5	3.39	7.8	94	3.45	4.43
6	4.95	7.9	88	5.25	6.31
7	2.89	8.6	90	2.36	3.23
8	6.40	9.1	100	7.09	8.69
9	6.01	7.3	89	6.49	7.28
10	6.94	8.8	89	7.36	9.07
11	5.86	8.2	93	5.46	6.69
12	8.35	9.6	97	6.58	8.75

Formulation of LP for unit *k*

Recall that the general statement of the Objective function in our DEA approach as,

$$\max E_k = u_1 O_{1k} + u_2 O_{2k} + \ldots u_M O_{Mk} \quad (7)$$

For example, for unit 1 we have,

$$\max : 5.98u_1 + 7.7u_2 + 92u_3$$

Note that these outputs do not share a common metric. Profit is in dollars while Satisfaction, and Cleanliness scores are outputs from some rating system. The fact that these measures use different metrics is not a major problem for a DEA analysis as long as each of the 12 units are all scored on the same basis. Again, the constraints are that the weights used (u_1, u_2, u_3, v_1, *and* v_2) cannot result in a unit having an efficiency level above 100%. Formally, we must have,

$$5.98u_1 + 7.7u_2 + 92u_3 - 4.74v_1 - 6.75v_2 \le 0,$$
$$7.18u_1 + 9.7u_2 + 99u_3 - 6.38v_1 - 7.42v_2 \le 0,$$
$$4.97u_1 + 9.3u_2 + 98u_3 - 5.04v_1 - 6.35v_2 \le 0,$$
$$5.32u_1 + 7.7u_2 + 87u_3 - 3.61v_1 - 6.34v_2 \le 0,$$
$$3.39u_1 + 7.8u_2 + 94u_3 - 3.45v_1 - 4.43v_2 \le 0,$$
$$4.95u_1 + 7.9u_2 + 88u_3 - 5.25v_1 - 6.31v_2 \le 0,$$
$$2.89u_1 + 8.6u_2 + 90u_3 - 2.36v_1 - 3.23v_2 \le 0,$$
$$6.40u_1 + 9.1u_2 + 100u_3 - 7.09v_1 - 8.69v_2 \le 0,$$

$$6.01u_1 + 7.3u_2 + 89u_3 - 6.49v_1 - 7.28v_2 \le 0,$$
$$6.94u_1 + 8.8u_2 + 89u_3 - 7.36v_1 - 9.07v_2 \le 0,$$
$$5.86u_1 + 8.2u_2 + 93u_3 - 5.46v_1 - 6.69v_2 \le 0,$$
$$8.35u_1 + 9.6u_2 + 97u_3 - 6.58v_1 - 8.75v_2 \le 0.$$

Next, recall that we are only able to write these efficiency constraints in this convenient form by scaling the input for the unit in question to equal 1. Thus our last constraint is,

$$4.74v_1 + 6.75v_2 = 1.$$

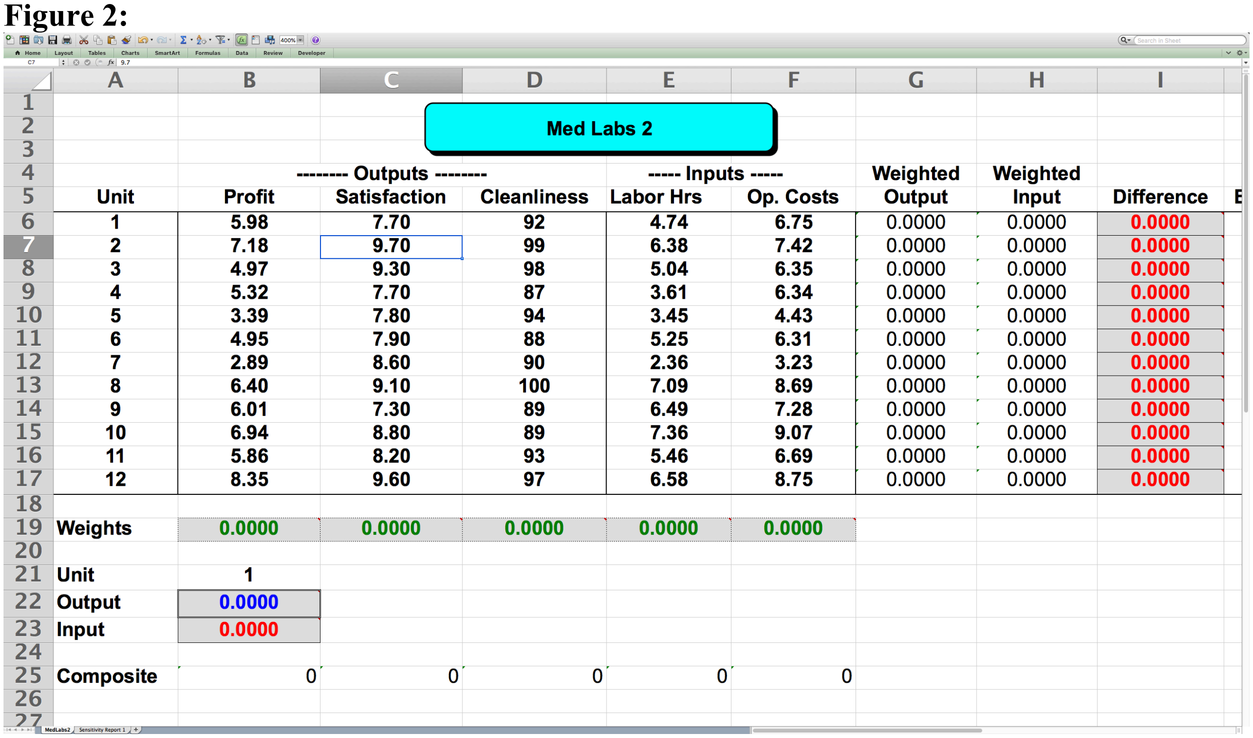
Don't, forget that this formulation only makes economic sense if all of the decision variables (u's and v's) are non-negative. In other words we must have,

$$u_1, u_2, u_3, v_1, v_2 \ge 0.$$

A convenient way to set this model up in spreadsheet form is displayed in Figure 2. (The Excel file containing this model is provided with this note.)

Several comments on this spreadsheet arrangement are useful. First the cells B19 through F19 are reserved to represent the weights for each of the input and output values. For unit 1 the outputs are in cells B6:D6 and the weights for these outputs are placed in cells B19:D19. Thus the formula in cell G6 is,

Figure 2:



	A	B	C	D	E	F	G	H	I
1									
2				Med Labs 2					
3									
4			-------- Outputs --------		----- Inputs -----		Weighted	Weighted	
5	Unit	Profit	Satisfaction	Cleanliness	Labor Hrs	Op. Costs	Output	Input	Difference
6	1	5.98	7.70	92	4.74	6.75	0.0000	0.0000	0.0000
7	2	7.18	9.70	99	6.38	7.42	0.0000	0.0000	0.0000
8	3	4.97	9.30	98	5.04	6.35	0.0000	0.0000	0.0000
9	4	5.32	7.70	87	3.61	6.34	0.0000	0.0000	0.0000
10	5	3.39	7.80	94	3.45	4.43	0.0000	0.0000	0.0000
11	6	4.95	7.90	88	5.25	6.31	0.0000	0.0000	0.0000
12	7	2.89	8.60	90	2.36	3.23	0.0000	0.0000	0.0000
13	8	6.40	9.10	100	7.09	8.69	0.0000	0.0000	0.0000
14	9	6.01	7.30	89	6.49	7.28	0.0000	0.0000	0.0000
15	10	6.94	8.80	89	7.36	9.07	0.0000	0.0000	0.0000
16	11	5.86	8.20	93	5.46	6.69	0.0000	0.0000	0.0000
17	12	8.35	9.60	97	6.58	8.75	0.0000	0.0000	0.0000
18									
19	Weights	0.0000	0.0000	0.0000	0.0000	0.0000			
20									
21	Unit	1							
22	Output	0.0000							
23	Input	0.0000							
24									
25	Composite	0	0	0	0	0			
26									
27									

=SUMPRODUCT(B6:D6,$B$19:$D$19)

Similarly, the weighted input for each unit is computed in column H as,

=SUMPRODUCT(E6:F6,$E$19:$F$19)

The sumproduct function in Excel is essentially a Dot-product. For example the sumproduct of (1,2,3) and (4,5,6) is (1 * 4) + (2 * 5) + (3 * 6) = 32. In other words this function multiplies the corresponding values in 2 lists and adds the results together.

The differences between the weighted outputs and inputs are computed in column I. Since efficiency can never be above 100% we know that the total weighted output must be no greater than the total weighted input. This is included in our formulation by adding the constraint that cells I6:I19 are less than or equal to 0.

We indicate which unit is being evaluated in Cell B21. Recall that when we evaluate unit *k* we need the objective function to reflect this. We also need the last constraint to reflect this fact as well. We can use another simple Excel function to facilitate this for all units to be evaluated.

Cell B22 contains a formula that returns the weighed output for this unit from the list of weighted outputs in column G. Maximizing this value is the objective. The formula in cell B22 is,

=INDEX(G6:G17,B21,1)

This function looks at the list of cells G6:G17 and takes the value in the row indicated in B21. Because the cell in B21 contains the number 1, this formula returns the value from the first row in the range, or G6. Similarly, if we look at unit 2, 3, 4 etc. this function will return the 2^{nd}, 3^{rd}, 4^{th}, entry etc. from this range.

We can use this same technique to make sure that we have the correct final constraint for the unit under consideration. Thus in Cell B23 we use,

=INDEX(H6:H17,B21,1)

Thus for whatever unit number is listed in cell B21, cell B22 represents the appropriate objective function to be maximized and cell B23 represents the weighted input that must be constrained to equal 1.

To solve this model we invoke Solver in Excel from the Tools menu and specify the target cell, variables cells, and constraints as shown in Figure 3. We include the non-negativity constraints by clicking on the box labeled "Make Unconstrained Variables Non-Negative." Finally, we use the dropdown menu labeled "Select a Solving Method" to choose "Simplex LP". This selection is important because this instructs Excel to use an algorithm that also creates a sensitivity report that we will use later.

The optimal solution for unit 1 is shown in Figure 4. The result shows that the efficiency score for unit 1 shown in Cell B22 is 0.9667 and is therefore slightly inefficient. Repeating this process for all 12 units generates efficiency scores of:

- Unit 1 – 0.9667
- Unit 2 – 1.000
- Unit 3 – 0.8345
- Unit 4 – 1.000
- Unit 5 – 0.8426
- Unit 6 – 0.8259
- Unit 7 – 1.000
- Unit 8 – 0.7720
- Unit 9 – 0.8572
- Unit 10 – 0.7958
- Unit 11 – 0.9188
- Unit 12 – 1.000

Figure 3:

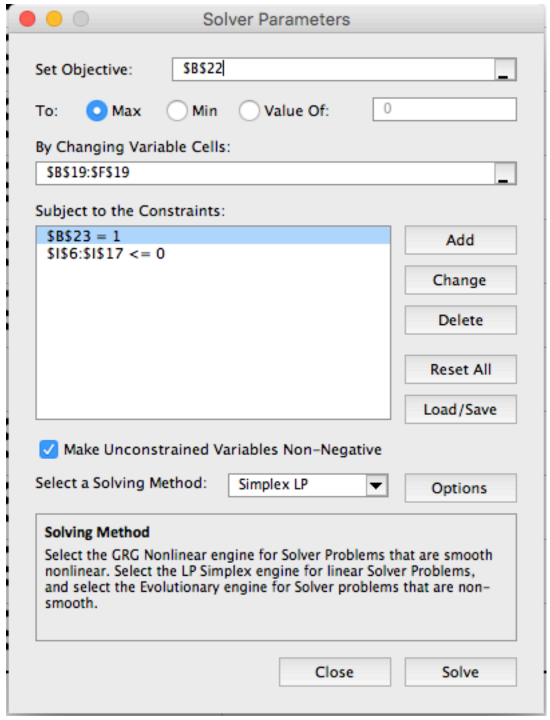
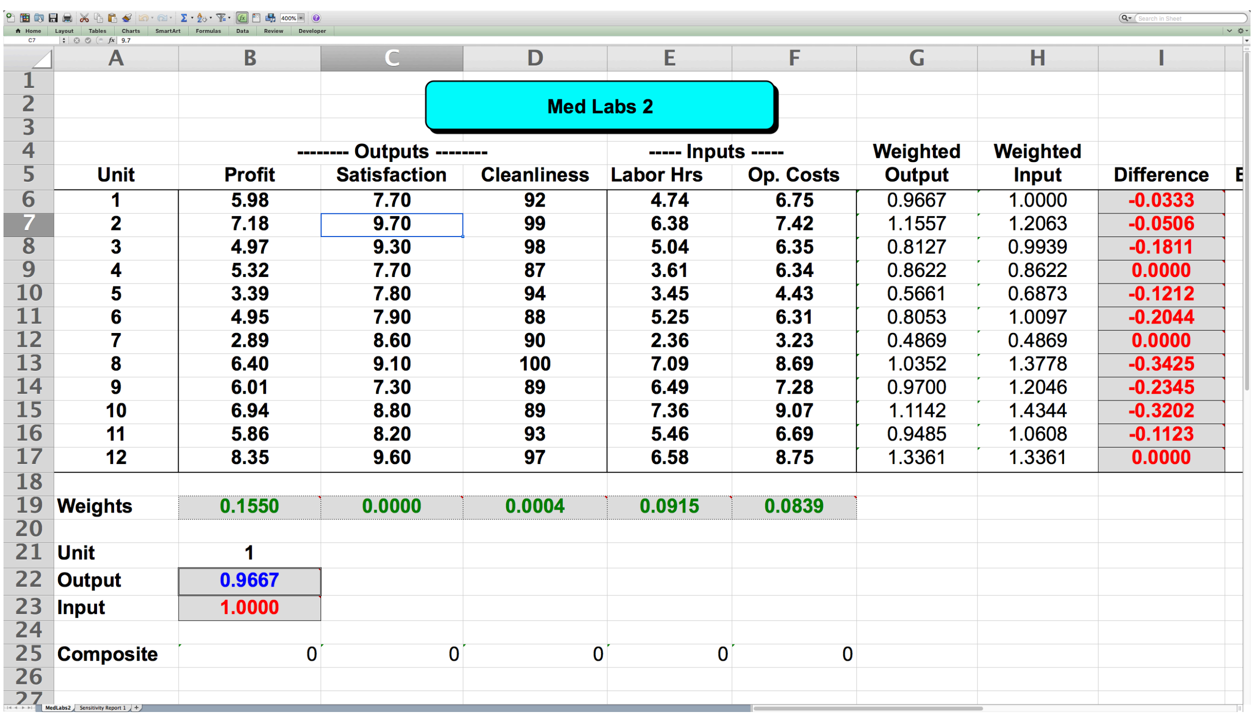


Figure 4:

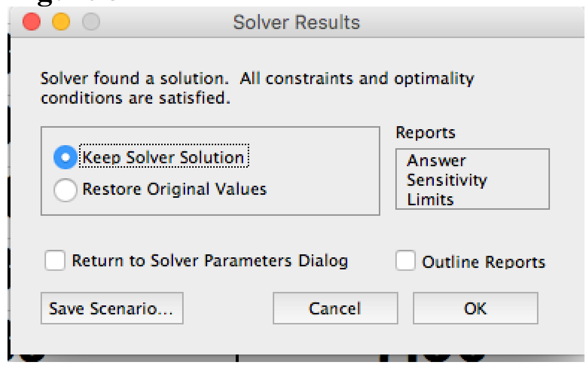


	A	B	C	D	E	F	G	H	I
1									
2			Med Labs 2						
3									
4		-------- Outputs --------			----- Inputs -----		Weighted	Weighted	
5	Unit	Profit	Satisfaction	Cleanliness	Labor Hrs	Op. Costs	Output	Input	Difference
6	1	5.98	7.70	92	4.74	6.75	0.9667	1.0000	-0.0333
7	2	7.18	9.70	99	6.38	7.42	1.1557	1.2063	-0.0506
8	3	4.97	9.30	98	5.04	6.35	0.8127	0.9939	-0.1811
9	4	5.32	7.70	87	3.61	6.34	0.8622	0.8622	0.0000
10	5	3.39	7.80	94	3.45	4.43	0.5661	0.6873	-0.1212
11	6	4.95	7.90	88	5.25	6.31	0.8053	1.0097	-0.2044
12	7	2.89	8.60	90	2.36	3.23	0.4869	0.4869	0.0000
13	8	6.40	9.10	100	7.09	8.69	1.0352	1.3778	-0.3425
14	9	6.01	7.30	89	6.49	7.28	0.9700	1.2046	-0.2345
15	10	6.94	8.80	89	7.36	9.07	1.1142	1.4344	-0.3202
16	11	5.86	8.20	93	5.46	6.69	0.9485	1.0608	-0.1123
17	12	8.35	9.60	97	6.58	8.75	1.3361	1.3361	0.0000
18									
19	Weights	0.1550	0.0000	0.0004	0.0915	0.0839			
20									
21	Unit	1							
22	Output	0.9667							
23	Input	1.0000							
24									
25	Composite	0	0	0	0	0			
26									
27									

Our results indicate that units 2, 4, 7, and 12 are operating at 100% efficiency (in the DEA sense) while the remaining units are operating less efficiently. We repeat that an efficiency rating of 100% does not necessarily mean that a unit is operating in the best possible way. It simply means that no linear combination of the other units in the set results in a composite unit that produces at least as much output using the same or less input.

We see that unit 1 has an efficiency score of 96.67% and is therefore somewhat inefficient. Recall that after running Solver to get an efficiency value for unit 1 we see the dialog box reproduced here as Figure 5.

Figure 5



When this box appears we can click on the "Sensitivity" button under "Reports". When this is done we get the report shown in Figure 6.

Figure 6

Microsoft Excel 14.6 Sensitivity Report
Worksheet: [Med_Labs_Example_1.xlsx]MedLabs2
Report Created: 5/23/2018 1:54:14 PM

Variable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
$B$19	Weights Profit	0.157644295	0	4.97	2.892852898	0.581789198
$C$19	Weights Satisfaction	0	-0.175446141	9.3	0.175446141	1E+30
$D$19	Weights Cleanliness	0.000520776	0	98	12.99284468	1.876369822
$E$19	Weights Labor Hrs	0.03044165	0	0	0.272386141	0.103933791
$F$19	Weights Op. Costs	0.133318754	0	0	0.130948328	0.34318492

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
$B$23	Input Profit	1	0.83452821	1	1E+30	1
$I$6	Difference	-0.053570716	0	0	1E+30	0.053570716
$I$7	Difference	0	0.329816495	0	0.0251409	0.050880644
$I$8	Difference	-0.16547179	0	0	1E+30	0.16547179
$I$9	Difference	-0.071160079	0	0	1E+30	0.071160079
$I$10	Difference	-0.11225865	0	0	1E+30	0.11225865
$I$11	Difference	-0.174892434	0	0	1E+30	0.174892434
$I$12	Difference	0	0.622431798	0	0.085425165	0.030166965
$I$13	Difference	-0.313370162	0	0	1E+30	0.313370162
$I$14	Difference	-0.174335542	0	0	1E+30	0.174335542
$I$15	Difference	-0.292851153	0	0	1E+30	0.292851153
$I$16	Difference	-0.085886118	0	0	1E+30	0.085886118
$I$17	Difference	0	0.096178403	0	0.045867829	0.032896325

We can look at the shadow prices that are above 0 in the Sensitivity report and deduce that a weighted average of 26.38% of unit 4 plus 28.15% of unit 7 plus 45.07% of unit 12 produces a hypothetical composite unit with outputs equal to those of unit 1 requiring less input.

We can use these shadow prices to create a theoretical unit, which is a composite of units 4, 7, and 12. This composite unit can be used as a target for unit 1. The composite values shown are simply weighted averages of the values for units 4, 7, and 12. Thus 26.38% * 5.32 + 28.15% * 2.89 + 45.07% * 8.35 yields a composite profit level of 5.98. A parallel calculation yields the Satisfaction, and Cleanliness levels along with Labor hours and Operating Costs for this hypothetical unit. It should not be surprising that some of these values are very close to those for unit 1 because that unit was very close to 100% efficiency. Comparing these composite values to those for unit 1 suggests an improvement strategy for that unit. The profit level is on par with the composite unit, as is its cleanliness score. A performance gap involving the outputs exists along the dimension of Customer Satisfaction (compare 8.8 to 7.7). Gaps exist involving both inputs. Ideally we would like to see unit 1 drive its labor hours down from 4.74 to 4.58 and to drive its Operating costs down from 6.75 to 6.53.

As an implementation issue, it is important to note that these results are not necessarily a result of bad management. There are many plausible explanations for this outcome. For example, the crew at unit 1 may be new and this lack of experience may be the explanation for the need for additional labor hours per unit of output. This extra labor requirement would also explain the higher operating costs. The major point is that an efficiency score below 100% implies that attention is warranted to uncover explanations. It is unwise to immediately jump to the conclusion that a managerial failure has occurred.

Figure 7:

Med Labs 2

	-------- Outputs --------			----- Inputs -----		Weighted	Weighted		DEA	Composite
Unit	Profit	Satisfaction	Cleanliness	Labor Hrs	Op. Costs	Output	Input	Difference	Efficiency	Weight
1	5.98	7.70	92	4.74	6.75	0.9667	1.0000	-0.0333	0.9670	0
2	7.18	9.70	99	6.38	7.42	1.1557	1.2063	-0.0506	1.0000	0.3298165
3	4.97	9.30	98	5.04	6.35	0.8127	0.9939	-0.1811	0.8345	0
4	5.32	7.70	87	3.61	6.34	0.8622	0.8622	0.0000	1.0000	0
5	3.39	7.80	94	3.45	4.43	0.5661	0.6873	-0.1212	0.8426	0
6	4.95	7.90	88	5.25	6.31	0.8053	1.0097	-0.2044	0.8259	0
7	2.89	8.60	90	2.36	3.23	0.4869	0.4869	0.0000	1.0000	0.6224318
8	6.40	9.10	100	7.09	8.69	1.0352	1.3778	-0.3425	0.7720	0
9	6.01	7.30	89	6.49	7.28	0.9700	1.2046	-0.2345	0.8572	0
10	6.94	8.80	89	7.36	9.07	1.1142	1.4344	-0.3202	0.7958	0
11	5.86	8.20	93	5.46	6.69	0.9485	1.0608	-0.1123	0.9188	0
12	8.35	9.60	97	6.58	8.75	1.3361	1.3361	0.0000	1.0000	0.0961784
Weights	0.1550	0.0000	0.0004	0.0915	0.0839					
Unit	1									
Output	0.9667									
Input	1.0000									
Composite	4.97	9.48	98	4.21	5.30					