A NON-LINEAR, MULTI-OBJECTIVE, AUTO-OPTIMIZATION MODEL FOR ALLOCATING RESOURCES TO CENTRALLY FUNDED HOSPITALS

Lawrence V. Fulton

Texas State University – San Marcos McCoy College of Business Administration Dept. of Quantitative Methods and Computer Information Systems 611 University Drive, San Marcos, TX 78666

lf25@txstate.edu 210-837-9977

Nathaniel D. Bastian

The Pennsylvania State University Center for Integrated Healthcare Delivery Systems Department of Industrial and Manufacturing Engineering 310 Leonhard Bldg, University Park, PA 16802 <u>nathaniel.bastian@fulbrightmail.org</u> 570-809-3619

Vivek P. Shah

Texas State University – San Marcos McCoy College of Business Administration Dept. of Quantitative Methods and Computer Information Systems 611 University Drive, San Marcos, TX 78666 vs01@txstate.edu

512-245-2049

ABSTRACT

This research explicates the similarities between data envelopment analysis (DEA) and multiple objective optimization (MOO) and proffers an original, hybrid, non-linear, multi-objective, resource allocation based optimization program that allows for the adjustment of resources (system inputs) either with or without decision-maker input. The motivation for this study is the necessity to balance resources among hospitals in large systems that are centrally controlled and funded, a task which has sometimes been based on historical funding levels with minor adjustments (e.g., centrally funded hospitals, such as those in the military). In these cases, inputs are fixed at certain levels and may only be adjusted within Decision-Making Units (DMUs). A mathematical formulation and example solutions based on both textbook and real data are provided.

Keywords: data envelopment analysis, multiple objective optimization, resource allocation, efficiency measurement, healthcare systems

BACKGROUND

Carlos Romero [6] illustrated that single objective, multi-objective and goal programming approaches are all special cases of distance models. An article written by Jorho, Korhonen, and Wallenius [5] expands Romero's idea by illustrating the similarity between Data Envelopment Analysis (DEA) and Multiple Objective Linear Programming (MOLP). This research explicates the similarities between DEA and multiple objective optimization (MOO) similar to Jorho et al, and proffers a hybrid, non-linear, multi-objective, resource allocation based optimization program that allows for the automatic adjustment of resources (system inputs) either with or without decision-maker input for governmental or military health systems with fixed inputs. A simple textbook formulation and solution of the hybrid model is included coupled with real-world analysis of U.S. Army military facility data. The motivation for this study is the necessity to balance resources among hospitals in large systems that are centrally controlled and funded while sustaining system output objectives.

MOTIVATING EXAMPLE

The Military Health System (MHS) attempts to balance health care cost, quality, and access across a \$4 billion enterprise. Doing so requires careful management of major system components. The driving postulate (based on the authors' previous research and involvement with decision makers) is that key leaders of this medical system wish to minimize inputs (primarily budgeted dollars and full-time equivalent healthcare providers) while maintaining outpatient weighted workload, inpatient weighted workload, prevention metrics, access metrics, and patient satisfaction. These components are linked in a complex, multi-objective fashion.

This paper first provides a discussion of multi-objective programming (MOOP) given the nature of the specified health care problem. Second, a discussion of a related method, DEA, is provided. After detailing these sets of methods, we detail the relationship between them, showing that they are nearly structurally identical.

After detailing DEA and MOOP, we discuss how neither provides a solution that re-directs slack to optimize total system performance. In other words, system efficiency (objectives) are subject to the assumption of fixed inputs. We therefore proffer a new, non-linear, multi-objective, autooptimization program that readjusts inputs automatically to maximize system efficiency. Such a program is useful for healthcare systems with fixed budgets and personnel authorizations (such as the U.S. Veteran's Hospital Administration). While non-linearity is inescapable, we demonstrate the formulation's effectiveness in solving a constant returns to scale (CRS, also known as CCR for Charnes, Cooper, and Rhodes) textbook healthcare problem as well as a variable returns to scale (VRS, also known as BCC for Banker, Charnes, and Cooper) real-world allocation problem associated with the U.S. Army hospital system. We begin with a discussion of multi-objective programming.

MULTI-OBJECTIVE PROGRAMMING

The motivating example consists of multiple competing objectives. While one might explore methods for modifying cost and/or production functions (e.g., Cobb-Douglas), the search here is

restricted to the field of optimization and begins with a discussion of the basic linear program (LP).

The typical linear program may be expressed in matrix notation as follows.

$$\begin{array}{l} \text{Max } \vec{dx} \\ \text{Subject to} \quad & A\vec{x} \leq \vec{b} \\ & \vec{x} \geq 0 \end{array} \tag{1}$$

Here, the objective function is composed of the $l \ x \ n$ coefficient vector \vec{d} and the $n \ x \ l$ decision variables \vec{x} . The constraint set is composed of the $m \ x \ n$ constraint coefficient matrix A along with our decision variables and the $m \ x \ l$ right hand side constraints b. Two objections to LP formulation are that linearity is often an over simplification of reality and that decision makers are rarely concerned with just one objective function, as in our motivating example (see French [4]). To address non-linearity, one might modify the formulation as follows.

Max
$$g(\vec{x})$$
 (2)
Subject to $\vec{x} \in X = \{\vec{x} \mid A\vec{x} \le \vec{b}, \vec{x} \ge 0\}$

This formulation provides a nonlinear function g(*), which accounts for real world complexity (and adds the same complexity to the solution algorithm). Still, the model does not consider multiple objective functions. The multi-objective optimization (MOO) follows.

V-Max
$$\vec{g}(\vec{x})$$

Subject to $\vec{x} \in X = \{\vec{x} \mid A\vec{x} \le \vec{b}, \vec{x} \ge 0\}$
(3)

Here, $\vec{g}(*)$ is a set of functions that define all objectives to be maximized (or minimized). The problem with this formulation is that conflicting goals may prevent the simultaneous optimization of all the objectives. Generally, a Pareto optimal solution set is sought such that for $\vec{y} \in X$ there exists no $g_i(\vec{y}) \ge g_i(\vec{x}) \forall \vec{x} \in X$ with strict inequality holding for at least one value of *i*. Unfortunately, finding the Pareto optimal still does not resolve the fundamental problem: which member of the efficient set does the decision maker choose? To answer this question, one might consider a value function as in the following formulation (see Cohen [2]).

V-Max
$$f(\vec{g}(\vec{x}))$$
 (4)
Subject to $\vec{x} \in X = {\vec{x} | A\vec{x} \le \vec{b}, \vec{x} \ge 0}.$

The value function f (*) is intended to monotonically increase with the decision maker's preference. Alternatively, the decision-maker may just explore efficient sets.

Returning to the motivating example, it was assumed that the senior medical decision-makers seek to minimize inputs while maintaining outputs constant. With an appropriately sufficient set of decision variables, one could attempt to devise a value function (as in f(*) above) based upon leaders' estimations of the importance of each item. In the case of this example, one might formulate the following set for optimization.

V-Max
$$f(g_i(\vec{x}))$$

Subject to $\vec{x} \in X = {\vec{x} | A\vec{x} \le \vec{b}, \vec{x} \ge 0}$
(5)

The objective function here is an $i \times i$ matrix coupled with an $i \times l$ set of decision variables for x. The constraint matrix A is of size $m \times i$.

Next assume that the *f* functions are ordered in a monotonic increasing fashion by preference. That is, $f(g_1^*)$ is less important to the decision-maker than $f(g_2^*)$ and so on. If this importance function is discrete, the referent-derived weighting system is similar to that of utility matrix weights. The inherent assumption is that the weighting system is developed consistently, i.e., that the decision-maker makes choices consistently in accordance with the value function *f*.

If one makes an assumption that f and g are linear, then the formulation is called the Multiple Objective Linear Program (MOLP) and looks familiar.

V-Max
$$\vec{v} = C\vec{x}$$

Subject to $\vec{x} \in X = \left\{ \vec{x} \mid A\vec{x} \le \vec{b}, \vec{x} \ge 0 \right\}$ (6)

Assume that there is interest in searching the non-dominated set of solutions. To do so with a linear f and g involves the projection of any point onto the set of non-dominated solutions. Wierzbicki [7] provides an achievement scalarizing function (ASF), which is capable of this projection given a feasible or infeasible starting point. The ASF will be discussed later, after an investigation of the structurally-related math programming technique of DEA.

DATA ENVELOPMENT ANALYSIS

DEA is a set of flexible, mathematical programming approaches for the assessment of efficiency, where efficiency is often defined as a linear combination of the weighted outputs divided by a linear combination of the weighted inputs as in the Charnes, Cooper, and Rhodes (CCR) model, which is a constant returns to scale (CRS) formulation [3]. Assume that an organization wishes to assess the relative efficiencies of some set of comparable subunits. (The subunits are called Decision Making Units or DMUs.) For each DMU, there is a vector of associated inputs and outputs of managerial interest. In this case, the manager is interested in either maximizing the outputs while not exceeding current levels of inputs (output oriented) or minimizing the inputs without reducing any of the outputs (input oriented). Using the hospital example, the inputs are budget, health care provider FTEs, and available beds, while the outputs are inpatient and outpatient weighted workload, a prevention metric, and patient satisfaction. In the case of DEA,

the manager assumes that the traditional definition of engineering efficiency (ratio of weighted outputs to weighted inputs) will result in an acceptable solution for technical efficiency. With these assumptions in place, one may formulate the following fractional programming problem that may be solved to determine technical efficiency, defined (for now) as the ratio of weighted outputs to weighted inputs, for each separate DMU [7, p. 23].

Max
$$\theta = \frac{\vec{u}^T \vec{y}_o}{\vec{v}^T \vec{x}_o}$$

Subject to:

$$\begin{aligned} & \frac{\vec{u}^T \vec{y}_z}{\vec{v}^T \vec{x}_z} \le 1, \forall z \\ & \vec{u} \ge 0 \\ & \vec{v} \ge 0 \end{aligned}$$

In this formulation, there is a vector of outputs (\vec{y}) , a vector of inputs (\vec{x}) , and z DMUs. Efficiency is designated as θ . The index *o* identifies the selected DMU for which an efficiency score will be generated. This mathematical program is run *z* times, once to determine the efficiency of each DMU. (While MOLP simultaneously solves multiple objective functions given a value function, DEA optimizes efficiency for an individual DMU.) The components of the vectors \vec{u} and \vec{v} are the weights to be determined for the outputs and inputs respectively. This model defines efficiency for the selected DMU as the weighted linear combination of its outputs divided by the weighted linear combination of its inputs, subject to the constraint that, for each DMU (including the one whose index *z* is *o*), the efficiency cannot exceed one. All weights are restricted to be nonnegative. This formulation is nonlinear; however, if one seeks to maximize the outputs while maintaining inputs constant, it is trivial to normalize the weighted inputs such that they equal one.

$$\vec{v}^T \vec{x}_o = 1 \tag{8}$$

Multiplying the numerator and denominator of the objective function as well as constraint (7) and finishing by adding (8) to the constraint set yields the following formulation.

Max
$$W = \vec{u}^T \vec{y}_o$$
 (9)
Subject to:

$$\vec{u}^T \vec{y}_z - \vec{v}^T \vec{x}_z \le 0, \forall z$$
$$\vec{v}^T \vec{x}_o = 1$$
$$\vec{u} \ge 0$$
$$\vec{v} \ge 0$$

(7)

For consistency with much of the literature, this formulation is considered the dual, so taking the "dual of the dual" provides the primal. (The primal allows for better comparison with Multiple Objective Programming.) In standard form, the primal follows.

$$\operatorname{Min} \theta - \varepsilon (1^{T} s^{+} - 1^{T} s^{-})$$
subject to:
$$X \overline{\lambda} - \theta \ \overline{x}_{o} + \overline{s}^{-} = 0$$

$$Y \overline{\lambda} - \overline{y}_{o} - \overline{s}^{+} = 0$$

$$\overline{\lambda}, \overline{s}^{*}, \varepsilon \ge 0$$
(10)

Here, the epsilon is an Archimedean element, and the slacks (s^*) reflect output shortages and input excesses. A DMU that has an efficiency score of one and a zero-slack solution (for all slacks) is technically efficient or Pareto-Koopmans efficient. As defined in Cooper, Seiford, and Tone [3], Pareto-Koopmans efficiency is attained only if it is impossible to improve any input or output without worsening some other input or output. In all other cases, it is possible to improve one or more of the inputs or outputs without worsening any other input or output.

Returning to our motivating example, one notes that the formulation of the DEA model will provide efficiency scores and slack information. The importance of any objective function is allowed to be a function of automatically generated weights. If a decision maker deems quality is more important than access, then the above formulation does not provide a weighting system (e.g., the f function).

Fortunately, there exist a variety of DEA based linear programs that assign weights to inputs and outputs based on importance of items. For example, Cooper et al. [3] provide a weighted slacks-based model (W-SBM) with decision maker weights applied. This model is similar to goal programming and is provided below (in fractional form) for reference.

$$Min \ \rho = \frac{1 - \frac{1}{m} \sum_{i=1}^{m} \frac{w_i s_i^-}{x_{io}}}{1 + \frac{1}{t} \sum_{r=1}^{t} \frac{w_r s_r^+}{y_{ro}}}$$
(11)
$$\sum_{i=1}^{m} w_i = m, \sum_{r=1}^{s} w_i = t$$

Here, the *w* are weights and the *s* are slacks (input excesses or output shortages). The inputs belong to *X* and the outputs belong to *Y*. Returning to our example, one can readily see the capability of a decision maker to provide weights for budget and personnel inputs (index *i*) as well as cost, quality, and access outputs (index *r*). See Cooper et al. [3, p. 105] for a complete discussion.

THE RELATIONSHIP BETWEEN DEA & MOOP

With this basic foundation in place, the question becomes, how does DEA relate to multiple objective programming and how might one leverage its methods to help the decision maker balance competing objective functions. The relationship between the two approaches provides the first point of discussion.

Jorho, Korhonen, and Wallenius (1996) [5] illustrated that DEA and MOLP are structurally related, as each might be formulated similar to the CCR output oriented model. One first notes the need to restrict MOLP to solutions existing within the set of non-dominated criterion vectors through the use of a achievement scalarizing function (ASF), a function that projects any feasible or infeasible point onto the dominated set.

Consider the following formulation of an ASF provided by Wiersbicki (1980) [7] and simplified by the authors.

$$Min(r) = \min\left\{ \max_{i \in P} \left(\frac{\vec{z}_i - \vec{c}_i \vec{x}}{\beta_i} \right) + \rho \sum_i z_i - \vec{c}_i \vec{x} \right\}$$

s.t. $\vec{x} \in X = \left\{ \vec{x} \mid A \vec{x} \le \vec{b}, \vec{x} \ge 0 \right\}$ (12)

where

 $\beta_i > 0 \equiv$ referent weights for each objective function

 $\rho > 0 \equiv$ an Archimedian element

 $i \equiv$ number of objective functions

 $z_i =$ the aspiration level for objective function (vector) *i*

 $\vec{c}_i \vec{x} \equiv$ the current objective function location of the solution for vector *i*

 $c \in C \equiv$ matrix of objective function coefficients

 $i \in P, P$ is the set of objective function vectors

Quite simply, one seeks to find \vec{x} that minimizes the largest deviation between the aspiration location (in objective function space) and our current location (in objective function space), i.e., $\min\left(\max_{i\in P}\left(\frac{\vec{z}_i - \vec{c}_i \vec{x}}{\beta_i}\right)\right)$, while ensuring that the "slacks" for all vectors are as small as possible, i.e., $\min \rho \sum_i z_i - \vec{c}_i \vec{x}$. This simplistic explanation provides the basis for the formulation of the

ASF.

Following [5], one can further simplify our objective function with a simple replacement.

$$\min(r) = \min\left\{\varepsilon + \rho \sum_{i} z_{i} - \vec{c}_{i} \vec{x}\right\}$$
(13)

Subject to

$$\varepsilon \ge \left(\frac{\vec{z}_i - \vec{c}_i \vec{x}}{\beta_i}\right), i = 1, 2...p =$$

$$\varepsilon \vec{\beta} \ge \vec{z}_i - \vec{c}_i \vec{x} =$$

$$C \vec{x} + \varepsilon \vec{\beta} - \vec{v} = \vec{z}$$
and
$$A \vec{x} \le \vec{b} =$$

$$A \vec{x} - \vec{b} - \vec{y}_o = \vec{y}_o$$

$$A \vec{x} - \vec{s}^+ = \vec{y}_o$$

$$\vec{x}, \vec{z} \ge 0$$

One should also note the following.

$$\varepsilon + \rho \sum_{i} z_{i} - \vec{c}_{i} \vec{x} = \varepsilon + \rho(\varepsilon \mathbf{1}^{T} \vec{\beta} - \mathbf{1}^{T} \vec{v}) = \varepsilon(\mathbf{1} + \rho \mathbf{1}^{T} \vec{\beta}) - \rho \mathbf{1}^{T} \vec{v}$$
(14)

Since $\rho l^T \vec{\beta}$ is a constant, it may be removed from the minimization, leaving $\varepsilon - \rho l^T \vec{v}$. Letting $C = X, A = Y, \vec{x} = \vec{\lambda}, \varepsilon = \theta, -\vec{\beta} = \vec{x}_o, -\vec{v} = \vec{s}^-$, the final formulation results, one which is structurally similar to that of the CCR Input oriented model.

$$Min r = \theta + \rho 1^{T} \vec{s}^{-}$$

$$Subject to:$$

$$X\vec{\lambda} - \theta \vec{x}_{a} + \vec{s}^{-} = \vec{z}$$
(15)

$$\begin{aligned} X\lambda &= 0X_o + S^- = Z \\ Y\lambda &= \vec{s}^+ = \vec{y}_o \\ \lambda, \vec{s}^+, \vec{s}^-, \rho \ge 0 \end{aligned}$$

In fact, placing the models side-by-side reveals few structural differences (modified from [5]).

CCR DEA Model	Reformulated Referent Point Model
$\operatorname{Min} r = \theta - \varepsilon (1^T s^+ - 1^T s^-)$	$Min \ r = \theta + \rho 1^T \vec{s}^-$
subject to:	subject to:

$X\vec{\lambda} - \theta \vec{x}_o + \vec{s}^- = \vec{0}$	$X\vec{\lambda} - heta \ \vec{x}_o + \vec{s}^- = \vec{z}$
$Y\vec{\lambda} - \vec{y}_o - \vec{s}^+ = \vec{0}$	$Y\vec{\lambda} - \vec{y}_o - \vec{s}^+ = 0$
$\vec{\lambda}, \vec{s}^*, \varepsilon \ge 0$	$ec{\lambda},ec{s}^+,ec{s}^-, ho\geq 0$

With this comparison in hand, one might turn to an example of MOOP, DEA, and a hybrid model for discussion.

MOTIVATING EXAMPLE COMPARISON

Let us return to the example regarding military hospitals in the MHS. The components of the hospitals' activities are intrinsically linked. A possible multiple objective formulation related to these components might be the following.

$$\begin{array}{l} Max f(g(x1|x2,x3,x4,x5,x6,x7)) \\ Max f(g(x2|x1,x3,x4,x5,x6,x7)) \\ Max f(g(x3|x2,x1,x4,x5,x6,x7)) \\ Max f(-g(x4|x2,x3,x1,x5,x6,x7)) \\ Max f(-g(x5|x2,x3,x4,x1,x6,x7)) \\ Max f(-g(x6|x2,x3,x4,x5,x1,x7)) \\ Max f(-g(x7|x2,x3,x4,x5,x1,x7)) \\ max f(-g(x7|x2,x3,x4,x5,x1,x7) \\ max f(-g(x7|x2,x3,x4,x5,x1,x7)) \\ max f(-g(x7|x2,x3,x4,x5,x1,x7)) \\ max$$

This model seeks to maximize outputs while minimizing inputs subject to an f function determined by decision makers. Alternatively, the original DEA program (CCR with input orientation for consistency) might be employed with a Y matrix consisting of output measures and an X matrix consisting of input measures. Of course, the DEA formulation generally seeks weights independent of a referent function; however, one could program the referent function as a series of constraints, as in the W-SBM model.

LIMITATIONS OF MOO AND DEA FOR FIXED INPUT SYSTEMS

Decision-makers generally seek to investigate how inputs / outputs might be adjusted to improve the objectives. The reduced cost information certainly provides a starting point for analysis, and system slack informs how one might be able to reduce inputs (for output-oriented models). The dual variables show the effects of a unit relaxation in constraint might have on the objective function. Finally, sensitivity analysis associated with adjustments in any portion of the model can help inform decision makers. In fixed input systems, it becomes necessary to improve system performance by adjusting inputs in the system subcomponents (DMUs). A multiobjective model that adjusted resources automatically across all facilities to achieve maximum system efficiency would (at a minimum) provide decision-support and insight for leaders interested in evaluating multiple objectives simultaneously. In this next section, we provide an alternative, multi-objective formulation that is based on a super-objective applied to traditional DEA analysis. This model assumes that the decision-maker would like to change inputs and outputs in order to have the resources necessary to achieve at least a minimum level of performance. We specify the formulation of the model, and provide a CRS textbook example as well as a VRS real-world example. A discussion of the model follows.

MULTI-OBJECTIVE AUTO-OPTIMIZATION MODEL

In the preceding discussion, we demonstrated that DEA and MOOP are related methods for evaluating multiple objective problems. In DEA, we noted that the weights are determined via optimization, while in MOOP, these weights are assigned. While both the MOOP and DEA formulations provide possible courses of action for decision makers, we propose a DEA-Based, Multi-Objective Auto-Optimization Model (AOM) for specific cases where one seeks to balance system components that might be interpreted as a performance ratio (not necessarily efficiency) as in the motivating example. Such a formulation should be able to identify inputs that might be manipulated to improve system performance over multiple outputs (objectives). Essentially, this formulation should be able to provide sensitivity analysis to advise decision makers how to optimally reallocate resources in order to attain the most efficient system possible. The next formulation applies to any multiple objective problems that involve fixed inputs (or possibly outputs) that are fixed en toto but can vary between DMUs. For example, we use a fixed budget large hospital organization (such as the military hospital system, the veterans' hospital system, and governmental hospital systems) that may reallocate resources among its facilities. A description of the model, its derivation, and an application follow. The definition of variables, sets, and data matrices follows.

Indices

 $o \equiv$ index of all *m* DMUs(hospitals) $j \equiv$ index for outputs $k \equiv$ index for inputs

Decision Variables

 $\partial_{ko} \equiv adjustments to each input k by DMUo with <math>\partial \in \Delta$

 $\alpha_{jo} \equiv$ weight for output *j* and DMU*o* with $\alpha \in A$

 $\lambda_{ko} \equiv \text{weight for input } k \text{ and } DMUo \text{ with } \lambda \in \Lambda$

 $r \equiv$ lower limit for efficiency score required for all DMUs

Data

 $x_{ko} \equiv \text{input } k \text{ for DMUo with } y \in Y$ $y_{io} \equiv \text{output } j \text{ for DMUo with } y \in Y$

$$Max \ z = \sum_{o} \sum_{j} a_{jo} y_{jo} \tag{17}$$

Subject to:

$$r \le \sum_{j} \alpha_{jo} y_{jo}, \forall o \tag{18}$$

$$\sum_{j} \alpha_{jo=v} y_{jo} - \sum_{k} \lambda_{ko=v} (x_{ko} + \delta_{ko}) \le 0, \forall o, v \in \{1, 2, ... N\}$$
(19)

$$\sum_{k} \lambda_{ko} (x_{ko} + \delta_{ko}) = 1, \forall o$$
⁽²⁰⁾

$$x_{ko} + \partial_{ko} \ge 0, \forall k, o \tag{21}$$

$$\sum_{o} \partial_{ko} = 0, \forall k \tag{22}$$

$$0 \le r \le 1$$

$$a_{jo} \ge 0 \forall j, o$$

$$\lambda_{ko} \ge 0 \forall k, o$$

$$\Delta \text{ free}$$
(23)

The objective function (17) seeks to optimize the sum of the efficiencies for all of the DMUs, which are the weighted outputs in this AOM model. (Note: subtracting a convexity variable and constraint converts this to a VRS model when also subtracted from the constraints associated with 19.) In (18), the weighted outputs are restricted to be strictly greater than a global efficiency variable r, which exists on [0, 1]. This constraint is important as one could imagine the objective function seeking to reduce the efficiency of one DMU to near zero in order to make the others nearer to one.

In (19), we force the sum of the weighted outputs to be less than or equal to the sum of the weighted inputs after adjusting them up or down by the amount necessary to achieve the highest sum of efficiency scores for each selected DMU (o=v). This constraint applies weights generated for each separate DMU analysis to all other DMUs inputs and outputs for relative efficiency comparison, just as is done in traditional DEA. This constraint makes the problem non-linear since the input weights are multiplied against the input changes.

In (20), we force the sum of the weighted and adjusted inputs to be equal to one. Doing so ensures that we will have efficiency scores for each DMU less than or equal to one. Again, this constraint is nonlinear.

In (21), we force each remaining input (after adjustment) for each DMU to be strictly greater than zero. Negative resources are not feasible.

The constraints in (22) require that any input adjustments sum to zero. We cannot grow resources for reallocation. Finally, the last set of constraints depicted in (23) are the usual bounds for the decision variables.

One might also include management constraints regarding the maximum movement of resources to increase flexibility and reflect management input into the system, similar to the f function provided by the MOO programming. Doing so would simply require bounds on the appropriate δ . These constraints would represent decision-maker input, similar to the development of the f function in multiple objective programming.

TEXTBOOK EXAMPLE

The solution to the AOM presented above provides the decision-maker recommendations regarding staffing of providers and allocation of funding such that all facilities achieve at least the efficiency associated with the r constraint. With this model formulation, there is a method for providing information regarding the adjustment of all inputs and outputs independent of or dependent upon decision-maker input.

Using Microsoft Excel's GRG Solver and GAMS CONOPT Solver, a simple, hospital-based textbook problem [1] was initially examined followed by a real-world example involving seven AMEDD hospitals with data from 2003. In the textbook example, seven different hospitals (DMUs) were initially evaluated using standard CRS DEA. The adjustable inputs for the hospitals included Full-Time Equivalents (FTEs), supply expenses in 1,000's, and available beds in 1,000's. Outputs include patient-days for those 65 and older in 1000's, patient-days for those under 65 in 1,000's, nurses trained, and interns trained. The data are shown in Table 1 below:

DMU	FTEs	Supply	Available	Patient	Patient	Nurses	Interns
		Expenses	Bed Days	Days >=65	Days <65	Trained	Trained
А	310.0	134.6	116.0	55.31	49.52	291	47
В	278.5	114.3	106.8	37.64	55.63	156	3
С	165.6	131.3	65.52	32.91	25.77	141	26
D	250.0	316.0	94.4	33.53	41.99	160	21
E	206.4	151.2	102.1	32.48	55.3	157	82
F	384.0	217.0	153.7	48.78	81.92	285	92
G	530.1	770.8	215	58.41	119.7	111	89

TABLE 1

Solving the problem using CCR DEA models results in all hospitals being efficient with the exception of Hospital D, which is 90.73% efficient. Reduced costs suggest that to enter the model, FTEs would need to be reduced by 12.16, expenses reduced by \$184. 63K, and the number of interns adjusted by 7.67. The reference set for DMU D includes hospitals A, B, and E (meaning that the dual values are non-zero).

From the sensitivity analysis, a conclusion might be to reduce resources for Hospital D. In this system, however, inputs are fixed. They may be spread across the hospital system but not cut (at

least in the short-term. This leads us to using the AOM formulation provided in (17) through (23), setting *r* (the minimum efficiency for any facility) to at least .95. Using the GRG nonlinear solver in Excel, a solution is reached within a few seconds, and the resultant analysis provides efficiency scores equal to one for all facilities. No solution could be better, although other alternate solutions to the same problem do exist. The input adjustment matrix provides recommendations for each DMU and each input that when re-implemented into the CCR DEA confirm all efficiency scores equal to one. The new values for the inputs follow in Table 2.

_	DMU	FTEs	Supply Expenses	Available Bed Days
	А	310.10	134.60	118.9
	В	278.50	114.30	108.99
	С	165.60	131.30	66.02
	D	250.00	316.00	87.40
	E	206.40	151.20	103.39
	F	384.00	217.00	153.82
	G	530.10	770.80	215.00

TABLE 2

The analysis suggests that by changing available bed days for facilities (which means adding or removing beds), the efficiency scores might be improved the most. The optimality attained for this problem is only one of several optimal solutions available. For example, using the CONOPT Solver in GAMS resulted in an alternate (but similar) solution set. Detailing multiple solution sets that result in maximizing the objective function is necessary to provide decision support.

REAL-WORLD EXAMPLE

With this simple example in hand, we move to the analysis of sixteen military medical facilities with inputs and outputs that were deemed important to decision-makers in evaluating efficiency. The data are from 2003 (as to be non-sensitive in nature), and the facilities were chosen from 24 facilities because they are largely homogenous. The inputs that could be manipulated included the funding stream (COST) and the FTEs (FTE). A non-discretionary input was the enrollment population supported (ENROLL). The outputs of interest included weighted workload metrics (inpatient relative weighted product known as RWP and outpatient relative value units known as RVU), a prevention metric (PREV), a satisfaction metric (SAT), and an access metric (ACC). These three metrics were scaled measures on [0,100]. The original data are shown in Table 3.

	ENROLL	FTE	COST	RWP	RVU	PREV	ACCESS	SAT
H1	14.81	7.13	56.66	7.05	112.21	83.28	70.55	73.19
H2	23.09	9.86	72.67	6.51	182.38	83.40	66.24	71.55
H3	68.40	17.66	163.99	21.74	372.06	78.89	57.29	63.02
H4	80.62	17.20	169.14	14.14	476.48	89.14	67.39	73.63

TABLE 3

H5	49.84	15.25	125.44	16.87	314.98	85.65	65.72	72.02
H6	38.13	13.04	130.23	10.41	229.08	84.82	65.61	69.87
H7	32.87	8.68	67.25	10.74	187.00	79.70	67.86	70.83
H8	12.74	6.34	53.16	7.07	85.10	84.60	67.49	73.67
H9	23.95	11.73	95.60	14.31	253.72	83.15	70.59	74.81
H10	14.93	6.42	52.37	0.96	76.53	89.44	65.40	69.85
H11	47.87	16.91	129.16	21.93	339.66	85.73	69.30	74.35
H12	31.50	8.81	71.98	3.01	153.56	82.32	60.92	69.81
H13	22.99	11.13	99.60	6.71	252.20	85.63	74.52	80.99
H14	31.39	12.73	92.53	14.87	298.59	83.97	70.48	75.62
H15	10.70	6.22	38.08	3.03	60.06	80.83	64.76	72.84
H16	63.40	14.71	114.29	14.86	327.31	80.24	62.89	68.53

For this more complex analysis, we ran variable returns to scale (VRS) DEA analysis, as such an analysis reasonably assumes that the production frontier is not necessarily linear. Assuming that enrollment is a nondiscretionary input, facilities with inefficiency scores less than 1.0 included H1 (.851), H2 (.928), H5 (.948), H7 (.779), H8 (.951), H9 (.850), H11 (.998), H13 (.959), H14 (.842), and H16 (.974).

Running the data in the AOM resulted in all facilities achieving efficiency scores of 1.0 by changing funding (COST) and FTEs as shown in Table 4. What is remarkable about these results is that the changes, while significant, are not so severe as to require side constraints. It is possible that the auto-optimization could recommend the elimination of FTEs or funding from a facility, which would require the use of side constraints. As an example, alternate formulations that we have run have included a maximum reduction of X% for any input, so as to prevent massive system changes. But with this real-world data, the solution set required only (relatively) minor changes to FTEs and budgeting.

	ORIGINAL COST	NEW COST	ORIGINAL FTE	NEW FTE
H1	56.66	53.41	7.13	7.94
H2	72.67	72.63	9.86	9.64
H3	163.99	163.98	17.66	16.35
H4	169.14	174.81	17.20	17.47
H5	125.44	125.43	15.25	13.33
H6	130.23	130.21	13.04	10.28
H7	67.25	64.98	8.68	10.42
H8	53.16	53.16	6.34	6.16
H9	95.60	95.59	11.73	11.62
H10	52.37	52.38	6.42	6.30
H11	129.16	129.15	16.91	17.00
H12	71.98	71.92	8.81	7.56
H13	99.60	99.61	11.13	10.55
H14	92.53	92.57	12.73	19.33

TABLE 4

H15	38.08	38.10	6.22	5.39
H16	114.29	114.25	14.71	14.48

Again, we note that multiple solutions are likely to be available for many problems. Investigating these multiple optimal solutions is something that is important in order to provide quality decision support.

SUMMARY AND CONCLUSIONS

By exploring the similarities between optimization methods for handling multiple objective problems, a related non-linear, multi-objective, resource allocation-based optimization program that allows for the adjustment of resources (system inputs) either with or without decision maker input was generated, programmed and solved on a representative data set. The utility for this type of decision support model to be employed in support of resource allocations for large, centrally funded hospital systems is self-evident. As demand for health services increases, the need for efficient allocation models based on competing objectives will become increasingly more important, and models similar to those proffered here will aid decision-makers' efforts.

This effort is not complete. Clearly, nonlinearity poses unique challenges for obtaining global optimality. The examples here are small and straightforward. That said, there is utility in being able to investigate reallocation strategies using automated methods. For instance, we are now beginning the use of multi-start genetic algorithms for gathering families of solutions that optimize over the problem space in order to provide decision makers arrays of possible solutions.

REFERENCES

[1] Anderson, D., Sweeney, D., Williams, T., Cam, J., and Martin. K. 2012. *An Introduction to Management Science*, Ohio: South-western (Cengage), p. 248.

[2] Cohen, J. 1978. Multiobjective Programming and Planning. Academic Press, NY.

[3] Cooper, W., Seifer, L., Tone, K. 2007. Data Envelopment Analysis, 2d Edition. Springer: New York.

[4] French, S. September 1984. Interactive Multi-Objective Programming: Its Aims, Applications and Demands. *The Journal of the Operational Research Society*, Vol. 35, No 9, pp. 827-834.

[5] Jorho, T., Korhonen, P. and Wallenius, J. 1998. Structural Comparison of Data Envelopment Analysis and Multiple Objective Linear Programming", Management Science, Vol. 44, No 7, pp. 962-970.

[6] Romero, C. March 1985. Multi-Objective and Goal-Programming Approaches as a Distance Function Model. *The Journal of the Operational Research Society*, Vol. 36, No. 3, pp. 249-251.

[7] Wierzbicki, A. 1980. The Use of Reference Objectives in Multiobjective Optimization. In G. Fandel and T. Gal, *Multiple Objective Decision Making Theory and Application.*, Springer-Verlag, New York.