

Wharton

Financial
Institutions
Center

*Projections Onto Efficient
Frontiers: Theoretical and
Computational Extensions to
D.E.A.*

by
Frances X. Frei
Patrick T. Harker

95-22-C

THE WHARTON FINANCIAL INSTITUTIONS CENTER

The Wharton Financial Institutions Center provides a multi-disciplinary research approach to the problems and opportunities facing the financial services industry in its search for competitive excellence. The Center's research focuses on the issues related to managing risk at the firm level as well as ways to improve productivity and performance.

The Center fosters the development of a community of faculty, visiting scholars and Ph.D. candidates whose research interests complement and support the mission of the Center. The Center works closely with industry executives and practitioners to ensure that its research is informed by the operating realities and competitive demands facing industry participants as they pursue competitive excellence.

Copies of the working papers summarized here are available from the Center. If you would like to learn more about the Center or become a member of our research community, please let us know of your interest.

Anthony M. Santomero
Director

*The Working Paper Series is made possible by a generous
grant from the Alfred P. Sloan Foundation*

Projections Onto Efficient Frontiers:
Theoretical and Computational Extensions to D.E.A. ¹

Revised: February 1998

Abstract: Data Envelopment Analysis (DEA) has been widely studied in the literature since its inception in 1978. The methodology behind the classical DEA, the oriented method, is to hold inputs (outputs) constant and to determine how much of an improvement in the output (input) dimensions is necessary in order to become efficient. This paper extends this methodology in two substantive ways. First, a method is developed that determines the least-norm projection from an inefficient DMU to the efficient frontier in both the input and output space simultaneously, and second, introduces the notion of the “observable” frontier and its subsequent projection. The observable frontier is the portion of the frontier that has been experienced by other DMUs (or convex combinations of such) and thus, the projection onto this portion of the frontier guarantees a recommendation that has already been demonstrated by an existing DMU or a convex combination of existing DMUs. A numerical example is used to illustrate the importance of these two methodological extensions.

Keywords: Data Envelopment Analysis, Productivity, Nonlinear Programming

Frances X. Frei is at the Simon School of Business, University of Rochester, Rochester, NY 14627, e-mail: frei@ssb.rochester.edu

Patrick T. Harker is at the Department of Operations and Information Management, University of Pennsylvania, Philadelphia, PA 19104-6366, e-mail: harker@opim.wharton.upenn.edu

1. Introduction

Managers are often faced with the task of evaluating relative performance of different people, teams, business units, etc. In cases where each of these performers has a set of common inputs that they utilize in order to produce a set of common outputs, Data Envelopment Analysis (DEA) has become one of the more popular tools for productivity analysis. This fact is evidenced by the more than eighty publications on DEA in 1994 and the first half of 1995 as well as the over thirty talks on DEA at the 1995 New Orleans INFORMS conference. DEA is especially useful when there are multiple inputs and outputs with different units of measure. That is, in instances when it is not desirable to translate each unit of measure to a common scale (say dollars), DEA is often the comparison method of choice. The purpose of these comparisons is to determine the best performers along with guidelines for improving the rest. Typically, these guidelines stem from benchmarking against a set of good performers in order for a poor performer to improve. Not surprisingly, the selection of the decision making units (DMUs) to benchmark against is a critical part of this analysis. The focus of this paper is in selecting the appropriate benchmarking set (or reference set) and understanding the implications of this choice for improving performance.

As will be described in detail in Section 2, when utilizing DEA to evaluate a set of (DMUs), an efficient frontier is created that determines which DMUs are performing well (efficient) and which are not (inefficient). The efficient set consists of those DMUs which are not inferior to any other DMU or a convex combination of any other DMUs along any dimension. Similarly, the inefficient DMUs are those that are dominated along at least one of the dimensions either by another DMU or a convex combination of other DMUs. There are several ways of applying the DEA methodology that all stem from the seminal paper of Charnes, Cooper and Rhodes (1978). These different methodologies typically give the same efficient set (with the exception of stochastic frontier analysis); that is, they build similar frontiers, but differ in their calculation of inefficiency. These calculations often relate to the projection of an inefficient DMU to the frontier but, oddly, none of the existing methods actually calculates the length of this projection. Rather, these methods typically calculate efficiency by moving solely along the input space or solely along the output space. Clearly, if the benchmarking set is a function of where the projection lands on the frontier, then each of these methods will potentially yield different results. If the purpose of the benchmarking set is to determine those efficient DMUs most like an inefficient DMU, then, unless the manager is constrained to change solely in either the input or output space, it is better to have the ability to move along all dimensions. The first contribution of this paper is to determine efficiency scores based on this least-norm projection to the frontier along both the input and output spaces simultaneously.

However, when determining the least-norm projection onto the frontier, it is possible to project to a point that is an extrapolation of existing points and thus, is theoretically possible but has not been observed by any of the existing DMUs (or convex combinations of existing DMUs). The second contribution of this paper is to introduce the *observable frontier* (that portion of the efficient frontier that is a convex combination of efficient DMUs and not an extrapolation of efficient DMUs) and develop a method to recompute the least-norm projection from the inefficient DMU to this observable frontier. As will be shown in the following sections, the recommendations based on the least-norm projection to the overall frontier as compared to those

based on the least-norm projection to the observable frontier are often significantly different, and lead to very different managerial conclusions.

The rest of the paper is organized as follows. The next section gives an overview of the related literature on DEA, Section 3 describes the shortest projection as well as the observable frontier, and Section 4 introduces an example which illustrates the vastly different implications that can come from each of these methods. Finally, Section 5 describes the shortcomings of the model developed in this paper as well as areas for further research.

2. Related Literature on Data Envelopment Analysis

Data Envelopment Analysis (DEA) is used to determine relative performance amidst multiple inputs and outputs. DEA was introduced by Charnes, Cooper, and Rhodes (1978) as a new method for measuring the efficiency of decision making units (DMUs). Since then, there have been over 400 articles that have used variations of DEA in analyzing performance (see Seiford 1993).

The original (input-oriented) DEA method determines the relative efficiency measure for a DMU by maximizing the ratio of weighted outputs to inputs subject to the condition that similar ratios for every DMU are not larger than one. The result of solving this problem is a set of efficiency scores less than or equal to one as well as a set of reference DMUs whose efficiency score is one using the same weights. This method has come to be known as the input-oriented method as its efficiency score is determined by holding outputs constant and assessing to what extent inputs would have to be improved (decreased) in order for a DMU to be considered efficient. An efficient DMU has no potential improvement, whereas inefficient DMUs have efficiency scores reflecting the potential improvement based on the performance of other DMUs. In order to determine the relative efficiency scores, a linear program (LP) must be run for each DMU. By using a linear objective function, the assumption is made that the efficient frontier is piecewise linear.

The DEA oriented-methods are described herein under the assumption of variable returns to scale. Slight modifications are necessary in order to accommodate constant returns to scale as described in Charnes et al (1982). The input-oriented method chooses, for a specific DMU, the multipliers that will maximize the ratio of outputs to inputs subject to the constraint that no DMU's output-to-input ratio can exceed one. Details of this method are readily available and can be found, among other places, in Fitzsimmons and Fitzsimmons (1998).

According to the early literature on DEA analysis, the efficiency score, which will be equal to one if a DMU is efficient and less than one if a DMU is inefficient, is the proportion by which all inputs must be reduced in order to become efficient. This is an important point: in the input-oriented method, not only are the outputs not changed but also, for a given DMU, each input is reduced by the same amount. Thus, for a particular inefficient DMU, in order to become efficient the same level of outputs would need to be maintained by using fewer inputs. Thus, the

projection onto the frontier is, in essence, calculated by reducing the input dimension until the DMU reaches the frontier. Of course, in the case where DMU n is efficient, the efficiency score will be one and the inputs would not have to change.

The reference set for a given DMU is the set of DMUs that are efficient using the optimal weighting for the given DMU. Mathematically, the reference set is the set of DMUs with non-zero dual prices for the DMU constraints in the LP. In addition, these dual prices will sum to one in the variable returns to scale case and will be less than or equal to one in the constant returns to scale case. (Actually, in the constant returns to scale case, the origin is implicitly in the reference set and thus, receives a portion of the weighting.)

In the case of the output-oriented BCC dual (BCC_D-O), where the objective is to minimize the ratio of inputs to outputs, the formulation is reversed. In this case, the result is a set of efficiency scores greater than or equal to one as well as a set of reference DMUs whose efficiency score is one using the same weights. Using this method, the efficiency score is determined by holding inputs constant and assessing to what extent outputs would have to be improved (increased) in order for a DMU to be considered efficient.

Many extensions have been made to the oriented methods described above, including multiplier weight flexibility (Dyson and Thanassoulis 1988), stochastic frontier (Sueyoshi 1994; Land, Knox Lovell, and Thore 1993), categorical outputs (Banker and Morey, 1986; Rousseau and Semple, 1993), and non-linear frontier estimation (Sengupta 1989; Charnes 1982). Sherman and Gold (1985) developed a method in which they determine the projection onto the frontier not by reducing each of the inputs by the same proportion as in the oriented methods but rather, by using the weighting scheme provided by the dual prices to calculate a composite point on the frontier. That is, rather than reducing every input by the efficiency score Z in order to determine the projection onto the frontier, they used the reference set of DMUs and their corresponding dual prices to determine their projection. However, this method does not project to the closest point on the frontier.

A third DEA model is the additive dual (ADD_D) (Charnes et al 1985; Ali and Seiford 1993). The optimal value yields an efficiency rating that measures the projection from a particular DMU to its associated facet of the frontier. The associated facet is the facet that is produced as a result of a particular DMU's LP. Again a single LP is required for each DMU, but rather than optimizing a ratio of inputs and outputs, the objective is to determine the coefficients of the hyperplane that will be closest to the current DMU without moving past any other DMU. A portion of this hyperplane becomes a facet of the efficient frontier. The result of solving this set of LPs is a projection of each DMU to its associated hyperplane, with a zero projection implying that the DMU is on the hyperplane and thus, on the frontier. Although this method was the first to actually deal with least-norm projections, the shortcoming is that the least-norm projection to the current facet is not necessarily (and in fact not usually) the least-norm projection to the entire frontier. The mathematical formulation for this method, which is similar to the other methods, is described in detail in (Ali and Seiford 1993).

The method described in the following section extends ADD_d by determining the least-norm projection from a DMU to the entire frontier. In addition, the concept of the observable portion

of the frontier is introduced which yields efficiency scores based on the projection from the portion of the frontier in which the associated scale has actually been realized (or is a convex combination of realized scale) rather than to a potential extrapolation of existing DMUs.

The additive dual to the multiplier DEA model is by no means the only paper that deals with least-norm projections to the frontier. For an excellent example of a least-norm projection method with an equi-proportionate deflation of all inputs (or an equi-proportionate inflation of all outputs), see Fare et al (1985). Banker and Morey (1986) provided the first paper in which a path to the frontier was different than the traditional Farrell radial contraction path.

In addition to the above mentioned methods, there is a large body of work done on sensitivity analysis and DEA which will not be reviewed in this paper. For references to this work, which involves taking minimum perturbations required to change a DMU from inefficient to efficient, see Charnes, Haag, Jaska, and Semple (1992), Rousseau and Semple (1995), and Charnes, Rousseau, and Semple (1996).

3. Efficiency Estimation with a Deterministic Frontier Method

Section 2 described the basic DEA or frontier estimation techniques. The intent of frontier estimation is to deduce empirically the production function in the form of an efficient frontier. That is, rather than knowing how to convert functionally inputs to outputs, these methods take the inputs and outputs as given, map out the best performers, and produce a relative notion of the efficiency of each. The problem with the existing methods is that they each measure efficiency in a conceptually suspect, albeit computationally effective, way. If the DMUs are plotted in their input/output space, then an efficient frontier that provides a tight envelope around all of the DMUs can be determined. The main function of this envelope is to get as close as possible to each DMU without passing by any others. A simple example of an efficient frontier (using variable returns to scale) is shown in Figure 1.

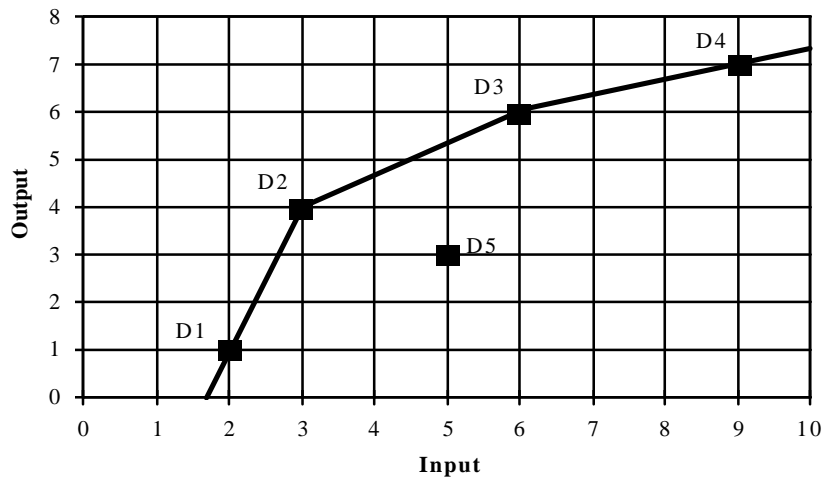


Figure 1. The Efficient Frontier

3.1 Shortest Projection

Each DMU along the frontier is considered efficient while those falling below the frontier, (e.g., D5) are considered inefficient. The method of determining the efficiency score for D5 varies according to the technique employed. Of the two classic methods, the input- or output-oriented methods, the efficiency score is determined, in effect, by determining the projection directly along the horizontal axis (holding outputs constant), or along the vertical axis (holding inputs constant). The method developed in this section determines the least-norm projection from an inefficient DMU to the frontier, in both the input and output space. In fact, one could develop a G-norm projection method (Ortega and Rheinbaldt 1970) wherein the input, output, and least two-norm projection are derived as special cases. For simplicity, this paper only describes the two-norm projection in order to illustrate the advantages of non-proportional projection methods.

The oriented methods require a series of LPs to be solved, one for each DMU. The method described in this section also requires the solution of a series of linear programs – the series represented by the multiplier dual to the additive DEA model – and then algebraically determines the least-norm projection. See Section 3.1.1 for a full description of the algorithm.

This least two-norm projection is illustrated in Figure 2 along with each of the oriented methods. Intuitively, the least-norm projection measure seems more appropriate than the other measures in that a DMU's benchmarking or reference set should be those efficient DMUs that offer some resemblance to the DMU. It is easy to imagine either of the oriented measures projecting to points on the frontier much farther away than the least-norm projection, thus yielding a reference set with presumably less in common than that of the closest projection. As can be seen, the efficient set of DMUs does not change with any of these methods, but the efficiency score, or relative projection, will be different in each case. These efficiency scores are useful in that they determine the relative inefficiency of a DMU; however, the real impact of this method is in the determination of the precise coordinates of the DMU, or convex combination of DMUs, against which a DMU can benchmark. For example, via the least-norm projection measure, D5 will reference D2 and D3 as benchmarks for efficient performance. D1 and D2 should only be the benchmarks under the case where one is unable to alter outputs. It is in this benchmarking or reference set that the dominance of this new method can be illustrated. Conceptually, it does not make sense to benchmark against firms that are potentially far away, except under circumstances when either the inputs or outputs are restricted¹.

¹ In banking, for example, there are circumstances when inputs may not be reduced (e.g., a policy of no layoffs combined with marketing efforts to increase the number of new accounts), outputs may not be augmented (e.g., given a fixed customer base, how much can FTEs be reduced?), or both can be changed simultaneously (e.g., a branch is inefficient and can undertake marketing to increase the number of accounts or layoff people to reduce FTEs)

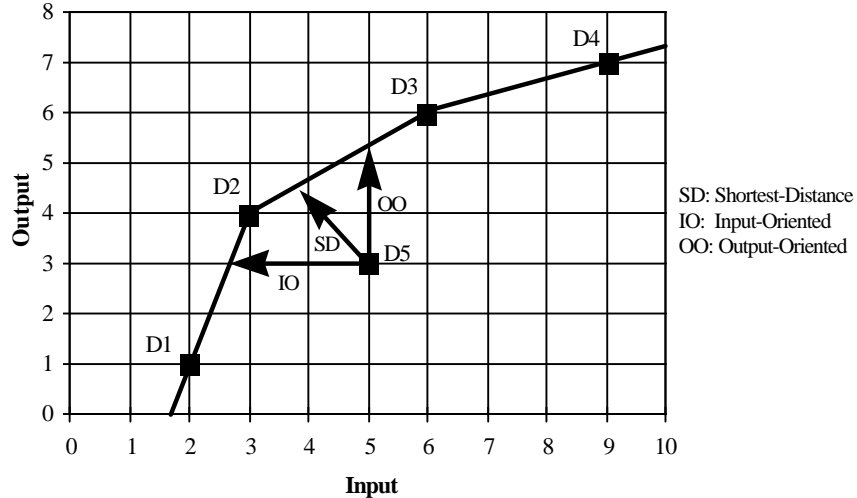


Figure 2. Single Input, Single Output Example

The mathematics of the method for determining this least-norm projection measure is described in detail in Section 3.1.1. However, it is important to note that this method requires the same number of LPs in its solution as each of the other methods. In addition, it is shown that the least-norm projection from an inefficient DMU is never at the intersection of two or more supporting hyperplanes of the efficient frontier. It is easy to see that in two dimensions, again referring to Figure 2, where no interior point will be closest to DMU 2. The two dimensional case is extended to n-dimensions in Section 3.1.1.

Although the oriented methods, in effect, project onto the frontier, they do not construct the frontier in their solution. This is significant because in understanding which facet of the frontier an inefficient DMU projects to, it is possible to determine the returns-to-scale of the projection. For example, the input-oriented projection in Figure 2 lands on the frontier at an increasing returns-to-scale segment while the least-norm projection and output-oriented each project onto a decreasing returns-to-scale segment. The returns-to-scale of a particular segment is directly attributable to the sign of the intercept of the supporting hyperplane. That is, if the sign is negative, the returns-to-scale is increasing, and if the sign is positive, the returns-to-scale is decreasing. If the intercept is at zero, there is constant returns-to-scale. Using the projections that actually construct the facets of the efficient frontier in their solution to the set of LPs, it is immediately apparent where a DMU projection lies and the relative returns-to-scale of that facet.

The significance of understanding the returns-to-scale is that in determining if the cost associated with improving performance is worthwhile, it is important to know the benchmarking environment of a DMU. Thus, the input-oriented projection shows that within the production possibility set there are other firms that are not only producing the same with less, but they are also enjoying increasing returns-to-scale (that is, the ability to get a relatively bigger jump in outputs than is required from the inputs). The least-norm projection measure shows that the closest portion of the frontier, one that involves the ability to change inputs and outputs simultaneously, will leave the DMU in a decreasing returns-to-scale environment. The managerial implications are obviously significant.

In addition, by using the oriented methods the projection onto the frontier is determined by reducing (increasing) the inputs (outputs) by an identical percentage based on the efficiency rating. Thus, if there are three inputs, as in the example in Section 4, then using the input-oriented approach not only requires that outputs are held constant, but also mandates that the reduction in inputs along each of the input dimensions is identical (and equal to the efficiency score). Sherman and Gold (1985) improve upon this by determining a projection based on the reference set of the DMU and the corresponding weight of each DMU in the reference set (the dual price of each DMU constraint as shown in Section 2). That is, the reference set composite DMU is used as the focus of what an inefficient DMU would look like if efficient. An example of this is shown in Table 2 in the next section.

Table 1 describes the relative efficiency scores from each of the methods for the DMUs in the simple example shown in Figure 1. In addition, Table 1 describes the reference set of each of the oriented methods as well as the weight of each reference DMU. For example, from Figure 2 we see that DMU 5 uses 5 units of input to produce 3 units of output. From Table 1, one sees that DMU 5 has an efficiency score of $8/15$. The classic input-oriented DEA method would deduce that the projection onto the frontier would multiply the inputs by the efficiency score in order to yield a projection of $5(8/15) = 2.67$. Thus, the projection of DMU 5 onto the frontier would land at the point $(2.67, 3)$. Sherman and Gold (1985) would determine the projection by determining the composite of the reference set DMUs 1 and 2: $0.33(2) + 0.67(3) = 2.67$. While these two methods result in the same projection in this 2-dimensional example, they are not guaranteed to do so in higher dimensions due to the likely adjustment required to account for slack.

Table 1. Efficiency Scores, Projection, and Reference Sets

DMU	Input Oriented Efficiency Score	Output Oriented Efficiency Score	Least-Norm Projection	Input Oriented Reference Set	Output Oriented Reference Set
1	1	1	0	1 (1.0)	1 (1.0)
2	1	1	0	2 (1.0)	2 (1.0)
3	1	1	0	3 (1.0)	3 (1.0)
4	1	1	0	4 (1.0)	4 (1.0)
5	0.53	1.78	1.94	1 (0.33)	2 (0.67) 2 (0.33) 3 (0.67)

3.1.1 Least-Norm Projection Algorithm

The fundamental problem with computing the least-norm projection (least two-norm) projection onto the efficient frontier is that, although the production set is convex, the frontier is non-convex. Thus, one must solve a least two-norm projection onto a piecewise-linear, non-convex surface. To overcome the problems caused by the non-convexity, the iterative procedure described below is used. Also, the following notion is used:

- μ vector of output weights
- v vector of input weights
- \mathbf{X} matrix of input values
- \mathbf{Y} matrix of output values

$$\sum_{r=1}^s \mu_r y_r - \sum_{i=1}^m v_i x_i + \mu_0 = 0 \quad \text{General equation for a hyperplane in } \mathfrak{R}^{m+s}$$

- Step 1. Solve a linear program (1) for each DMU in order to create the supporting hyperplanes $\mathbf{H} = \{H_i\}$ of the efficient frontier. Classify all hyperplanes as either unique hyperplanes $\mathbf{U} = \{U_i\}$ or non-unique hyperplanes. The unique hyperplanes are those hyperplanes generated by an LP with no multiple optima. The non-unique hyperplanes are those generated by LPs with multiple optima. Only the unique hyperplanes need to be considered going forward, as the non-unique hyperplanes will lie outside of the convex production set, except at a vertex of the production set. However, by Proposition 1, we know that the shortest distance to the frontier will never be at vertex of the production set, and thus all hyperplanes that are generated by LPs with multiple optima, can be ignored when calculating the shortest projection in Step 2.
- Step 2. Categorize every DMU as either efficient or inefficient. Efficient DMUs will be those DMUs that have zero as the optimal value of (1).

For each inefficient DMU²:

- Step 2. Calculate the least-norm projection d_i and the location of the projection algebraically to each unique hyperplane U_i .³
- Step 3. Compute $d^* = \min\{ d_i \}$, the least-norm projection to all unique hyperplanes \mathbf{U} .

² It is not necessary to calculate the shortest projection from efficient DMUs to the frontier as they are, by definition, on the frontier and thus their projection will be to themselves.

³ In the event of a tie, both projections are reported.

The algorithm described herein requires the solution of an LP for each DMU in order to determine the equations of the supporting hyperplanes on the efficient frontier, as described in Ali and Seiford (1993). Assuming there are n DMUs, m inputs, and s outputs, the general form for DMU $i=1, \dots, n$, is:

$$\begin{aligned}
& \max_{\mu, \nu, u_0} w_i = \mu^T \mathbf{Y}_i - \nu^T \mathbf{X}_i - u_0 \\
& \text{subject to} \\
& \mu^T \mathbf{Y}_i - \nu^T \mathbf{X}_i - u_0 \bar{\mathbf{1}} \leq 0 \\
& -\mu^T \leq \bar{\mathbf{1}} \\
& -\nu^T \leq \bar{\mathbf{1}} \\
& u_0 \text{ free}
\end{aligned} \tag{1}$$

where $\mathbf{Y} \in \mathfrak{R}^{s \times n}$ is matrix of outputs, $\mathbf{X} \in \mathfrak{R}^{m \times n}$ matrix of inputs, \mathbf{Y}_i is the column vector of outputs for DMU i , \mathbf{X}_i is the column vector of inputs for DMU i .

The solution to these n linear programs determines the efficient frontier. In order to find the closest point on the frontier to each DMU, one must compute a projection of the point $\langle \mathbf{X}_i, \mathbf{Y}_i \rangle$ onto each supporting hyperplane of the efficient frontier. Given a hyperplane defined by $(\mu^T \mathbf{Y}_i - \nu^T \mathbf{X}_i + u_0 = 0)$, one must solve the following nonlinear program to compute the least two-norm projection of $\langle \mathbf{X}_i, \mathbf{Y}_i \rangle$ onto this hyperplane.

$$\begin{aligned}
& \min_{\bar{\mathbf{X}}_i, \bar{\mathbf{Y}}_i} \sqrt{\|\mathbf{Y}_i - \bar{\mathbf{Y}}_i\|^2 + \|\mathbf{X}_i - \bar{\mathbf{X}}_i\|^2} \\
& \text{subject to} \\
& \mu^T \bar{\mathbf{Y}}_i - \nu^T \bar{\mathbf{X}}_i + u_0 = 0
\end{aligned} \tag{2}$$

This problem can be solved algebraically to produce the projection $\langle \bar{\mathbf{X}}_i, \bar{\mathbf{Y}}_i \rangle$ and the least-norm projection D_{1HP} :

$$\begin{aligned}
\bar{\mathbf{Y}}_i &= \mathbf{Y}_i - \frac{\mu(\mu^T \mathbf{Y}_i - \nu^T \mathbf{X}_i - u_0)}{\mu^T \mu + \nu^T \nu} \\
\bar{\mathbf{X}}_i &= \mathbf{X}_i + \frac{\nu(\mu^T \mathbf{Y}_i - \nu^T \mathbf{X}_i - u_0)}{\mu^T \mu + \nu^T \nu} \\
D_{1HP} &= \sqrt{\frac{(\mu^T \mathbf{Y}_i - \nu^T \mathbf{X}_i - u_0)^2}{\mu^T \mu + \nu^T \nu}} = \frac{|\mu^T \mathbf{Y}_i - \nu^T \mathbf{X}_i - u_0|}{\sqrt{\mu^T \mu + \nu^T \nu}}
\end{aligned} \tag{3}$$

Thus, the least-norm projection from each DMU to a given supporting hyperplane is known in closed form. By computing the least-norm projection for each hyperplane, one can then compute the minimum projection overall in Step 3 of the algorithm.

The only problem that could occur with the least-norm projection algorithm stated above is if the least-norm projection to the frontier occurs at the intersection of one or more supporting hyperplanes. This is a problem because the algorithm requires identifying a single hyperplane. The following proposition shows that this condition will never occur:

Proposition 1

The least two-norm projection of any point $(\mathbf{X}_i, \mathbf{Y}_i)$ onto the efficient frontier will never occur at the intersection of two or more hyperplanes; i.e., the projection will always be an orthogonal projection onto a single hyperplane H_i .

Proof:

First, note that the production set generated by the iterative solution of (2) will be convex. Second, assume that the projection occurs at the intersection of two hyperplanes H_i and H_j ; the extension to greater than two hyperplanes is straightforward. In this case, the projection is the solution of:

$$\begin{aligned} \min \text{OFV} &= \|\mathbf{Y}_i - \bar{\mathbf{Y}}_i\|^2 + \|\mathbf{X}_i - \bar{\mathbf{X}}_i\|^2 \\ \text{subject to} & \\ \mu_i^T \bar{\mathbf{Y}}_i - v_i^T \bar{\mathbf{X}}_i - u_{0_i} &= 0 \\ \mu_j^T \bar{\mathbf{Y}}_i - v_j^T \bar{\mathbf{X}}_i - u_{0_j} &= 0 \end{aligned} \tag{4}$$

The solution to (4) yields values of $(\bar{\mathbf{X}}_i, \bar{\mathbf{Y}}_i)$ as follows:

$$\begin{aligned} \bar{\mathbf{Y}}_i &= -\frac{\gamma_i \mu_i + \gamma_j \mu_j}{2} \\ \bar{\mathbf{Y}}_i &= \frac{u_{0_i} v_j - u_{0_j} v_i}{\mu_i v_j - \mu_j v_i} \\ \bar{\mathbf{X}}_i &= \mathbf{X}_i + \frac{\gamma_i v_i + \gamma_j v_j}{2} \\ \bar{\mathbf{X}}_i &= \frac{u_{0_i} c_i \mu_j - u_{0_j} \mu_i}{\mu_i v_j - \mu_j v_i} \end{aligned}$$

Finally, substitute $(\bar{\mathbf{X}}_i, \bar{\mathbf{Y}}_i)$ into the objective function of (4) in order to determine the least-norm projection, $D_{2\text{HPs}}$, from $(\mathbf{X}_i, \mathbf{Y}_i)$ to two hyperplanes:

$$\begin{aligned}
\bar{\mathbf{Y}}_i &= \mathbf{Y}_i - \frac{\gamma_i \boldsymbol{\mu}_i + \gamma_j \boldsymbol{\mu}_j}{2} \\
\bar{\mathbf{Y}}_i &= \frac{u_{0_i} \mathbf{v}_j - u_{0_j} \mathbf{v}_i}{\boldsymbol{\mu}_i \mathbf{v}_j - \boldsymbol{\mu}_j \mathbf{v}_i} \\
\bar{\mathbf{X}}_i &= \mathbf{X}_i + \frac{\gamma_i \mathbf{v}_i + \gamma_j \mathbf{v}_j}{2} \\
\bar{\mathbf{X}}_i &= \frac{u_{0_i} \boldsymbol{\mu}_j - u_{0_j} \boldsymbol{\mu}_i}{\boldsymbol{\mu}_i \mathbf{v}_j - \boldsymbol{\mu}_j \mathbf{v}_i} \\
OFV &= \|\mathbf{Y}_i - \bar{\mathbf{Y}}_i\|^2 + \|\mathbf{X}_i - \bar{\mathbf{X}}_i\|^2 \\
OFV &= \left\| \mathbf{Y}_i - \frac{u_{0_i} \mathbf{v}_j - u_{0_j} \mathbf{v}_i}{\boldsymbol{\mu}_i \mathbf{v}_j - \boldsymbol{\mu}_j \mathbf{v}_i} \right\|^2 + \left\| \mathbf{X}_i - \frac{u_{0_i} \boldsymbol{\mu}_j - u_{0_j} \boldsymbol{\mu}_i}{\boldsymbol{\mu}_i \mathbf{v}_j - \boldsymbol{\mu}_j \mathbf{v}_i} \right\|^2 \\
D_{2\text{HPs}} &= \sqrt{\left\| \mathbf{Y}_i - \frac{u_{0_i} \mathbf{v}_j - u_{0_j} \mathbf{v}_i}{\boldsymbol{\mu}_i \mathbf{v}_j - \boldsymbol{\mu}_j \mathbf{v}_i} \right\|^2 + \left\| \mathbf{X}_i - \frac{u_{0_i} \boldsymbol{\mu}_j - u_{0_j} \boldsymbol{\mu}_i}{\boldsymbol{\mu}_i \mathbf{v}_j - \boldsymbol{\mu}_j \mathbf{v}_i} \right\|^2} \quad (5)
\end{aligned}$$

In order to show that the shortest projection is never at the intersection of two hyperplanes, it need only be shown that the distance from $(\mathbf{X}_i, \mathbf{Y}_i)$ to two intersecting hyperplanes (5) is always greater than the distance from $(\mathbf{X}_i, \mathbf{Y}_i)$ to one hyperplane, $D_{1\text{HP}}$, given in (3). In order to show this, assume that the distance from $(\mathbf{X}_i, \mathbf{Y}_i)$ to hyperplane i is shorter than the distance from $(\mathbf{X}_i, \mathbf{Y}_i)$ to hyperplane j. Thus, it is only necessary to show that the distance to the intersection is greater than the distance to hyperplane i:

$$\begin{aligned}
\text{Distance}_{2\text{HPs}} - \text{Distance}_{1\text{HP}} &= \left\| \mathbf{Y}_i - \frac{c_i \mathbf{v}_j - c_j \mathbf{v}_i}{u_i \mathbf{v}_j - u_j \mathbf{v}_i} \right\|^2 + \left\| \mathbf{X}_i - \frac{c_i \mathbf{u}_j - c_j \mathbf{u}_i}{u_i \mathbf{v}_j - u_j \mathbf{v}_i} \right\|^2 - \frac{|\mathbf{u}_i^T \mathbf{Y}_i - \mathbf{v}_i^T \mathbf{X}_i - c_i|^2}{\mathbf{u}_i^T \mathbf{u}_i + \mathbf{v}_i^T \mathbf{v}_i} \\
&= \frac{\theta [u_i^T \mathbf{Y}_i |v_i^T v_j| - v_i^T \mathbf{X}_i |u_i^T u_j| - c_i |u_i^T u_j + v_i^T v_j|] - [u_i^T \mathbf{Y}_j |v_i^T v_i| - v_j^T \mathbf{X}_j |u_i^T u_i| - c_j |u_i^T u_i + v_i^T v_i|]^2}{|u_i^T v_j - u_j^T v_i|^2 |u_i^T u_i + v_i^T v_i|} > 0
\end{aligned}$$

□

3.2 A Small Numerical Example

Revisiting the example of Section 3.1, we demonstrate the use of the above algorithm in completing the shortest projection.

Step 1. Solve a linear program for each DMU in order to create the supporting hyperplanes of the efficient frontier. The LP for DMU D5 (see Section 3.1.2) is presented below with the results of all 5 LPs presented in Table 2. Note that only the objective function changes for each LP. We have drawn all five supporting hyperplanes (lines) in Figure 3. Note that the hyperplanes generated from D2 and D5 are dotted lines, while the other three are solid lines. The dotted lines are meant to represent the hyperplanes that are not unique due to multiple optima⁴. What should be clear from Figure 3 is that any hyperplane which is not unique (or not a frontier-defining hyperplane) will lie outside of the convex production set, except at a vertex of the production set. However, by Proposition 1, we know that the shortest distance to the frontier will never be at vertex of the production set, and thus all hyperplanes that are generated by LPs with multiple optima can be ignored when calculating the shortest projection in Step 2.

$$\begin{aligned}
 & \max w_0 = 3\mu_1 - 5\nu_1 - u_0 \\
 & \mu, \nu, u_0 \\
 & \text{subject to} \\
 & \mu_1 - 2\nu_1 - u_0 \leq 0 \\
 & 4\mu_1 - 3\nu_1 - u_0 \leq 0 \\
 & 6\mu_1 - 6\nu_1 - u_0 \leq 0 \\
 & 7\mu_1 - 9\nu_1 - u_0 \leq 0 \\
 & 3\mu_1 - 5\nu_1 - u_0 \leq 0 \\
 & -\mu_1 \leq 1 \\
 & -\nu_1 \leq 1 \\
 & u_0 \text{ free}
 \end{aligned}$$

Table 2. Hyperplanes generated from linear programs in Step 1

DMU	Original Data		Hyperplane generated from LP			Multiple optima	Efficient DMU
	output	input	μ_1	ν_1	μ_0		
D1	1	2	1	3	-5	no	yes
D2	4	3	1	1.33	0	yes	yes
D3	6	6	1.5	1	3	no	yes
D4	7	9	3	1	12	no	yes
D5	3	5	1	1	1	yes	no

⁴ It should also be clear from Figure 3 that all non-unique hyperplanes are convex combinations of unique hyperplanes and as such will be beyond the production set, except where they overlap with the unique hyperplanes. Thus, checking for the shortest projection to the unique hyperplanes is sufficient.

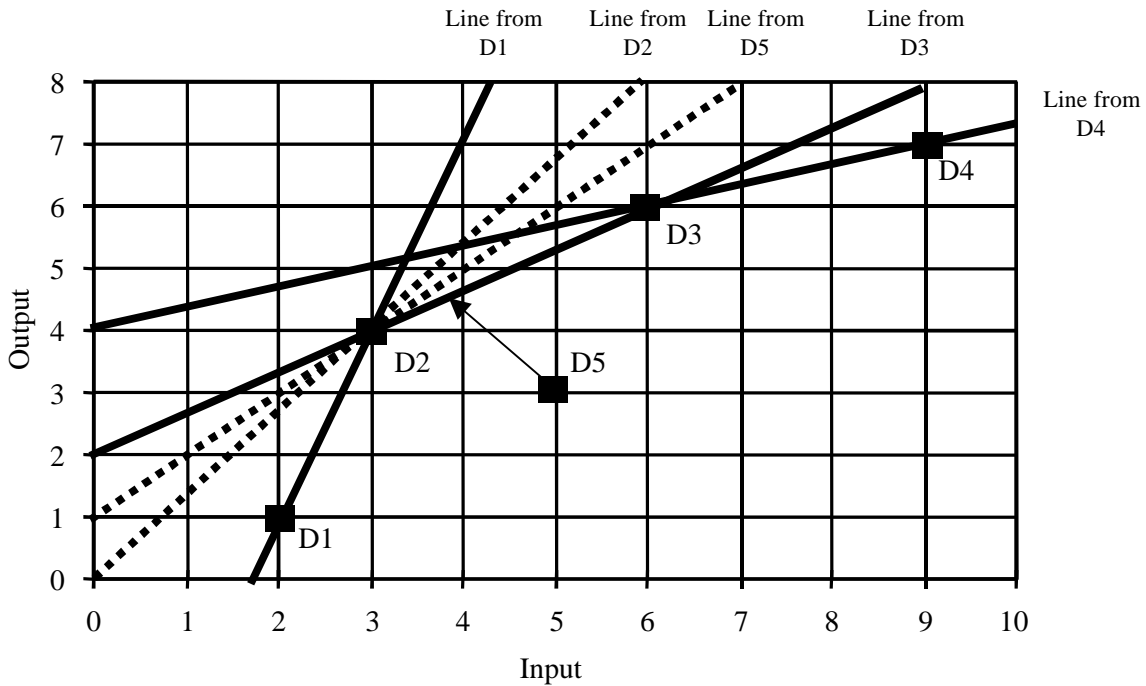


Figure 3. Hyperplanes from Small Numerical Example

For each inefficient DMU:

Step 2. Calculate the least-norm projection d_i and the location of the projection algebraically for each unique hyperplane U_i .

For the first four DMUs, the projection would be onto itself, but for the inefficient DMU D5, it is necessary to calculate the least norm projection to each of the three unique hyperplanes (see Table 3).

Table 3. Shortest distance from DMU D5 to each unique hyperplane in Step 1

Inefficient DMU	Projection to Unique Hyperplane	Output projection	Input projection	Shortest distance
D5	U_1	3.7	2.9	2.21
	U_3	4.62	3.92	1.94
	U_4	5.4	4.2	2.53

Step 3. Compute $d_* = \min\{ d_i \}$, the least-norm projection to all unique hyperplanes U .

As can be seen from the results of the three projections reported in Table 3, the shortest projection from D5 is to U_3 . The projection is to the point (3.92, 4.62) as shown in Figure 3.

The Observable Frontier

In utilizing projections from an inefficient DMU onto the efficient frontier, it is important to understand the implications of where the projection lands. That is, it may well be that the projection lands beyond any existing DMU or any convex combination of an existing DMU. If the efficient frontier is split into two sections, one that represents either an observable or convex combination of observed input-output combinations, and another that represents extrapolations of DMUs beyond those defined in the first section, then we have what we determine as the *observable* portion and *non-observable* portion of the frontier, respectively. There are many instances in which it is not practical to have a benchmark that extends beyond the scale observed by any existing DMUs (that is, to a *non-observable* point), and thus, independent projections onto the observable frontier are necessary. To illustrate this, Figure 3 duplicates the two-dimensional example from above with the observable portion of the frontier represented by a thicker line than the non-observable portion. In addition, another inefficient DMU has been added in order to illustrate that this DMUs shortest projection onto the entire frontier lands at a non-observable point. That is, the projected output for D6 is 7.2 whereas the largest observed output, that of D4, is only 7. There may very well be instances in which the shortest projection onto the observable frontier, in this case directly onto D4, is the best solution. Thus, it is necessary to report both the overall efficiency score as well as the observable efficiency score.

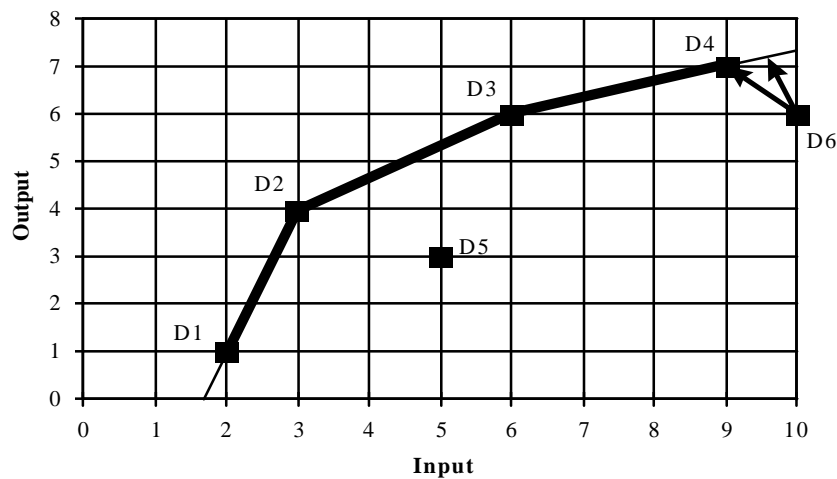


Figure 4. The *Observable Frontier*

3.2.1 Observable Frontier Algorithm

The following algorithm is used to compute the projection to the observable frontier:

- Step 1. Repeat steps 2 - 4 of the shortest project algorithm for each unique hyperplane U_i that makes up a facet of the efficient frontier.
- Step 2. Determine all DMUs that are on the supporting hyperplane H_i ; these DMUs are the reference set for this hyperplane.

- Step 3. Determine the convex hull of the reference set.
- Step 4. Determine the least-norm projection from each inefficient DMU to the convex hull; call this distance D_i . In order to determine the least-norm projection from a DMU to the convex hull, one must solve the following non-linear program (6). In this problem, I represents the set of DMUs in the reference set of H_i :

$$\begin{aligned}
 & \min \left\| \begin{pmatrix} \bar{Y}_i \\ \bar{X}_i \end{pmatrix} - \begin{pmatrix} \bar{Y}_H \\ \bar{X}_H \end{pmatrix} \right\|^2 \\
 & \text{subject to} \\
 & \begin{pmatrix} \bar{Y}_H \\ \bar{X}_H \end{pmatrix} \in S \\
 & S = \left\{ \sum_{i \in I} \lambda_i \begin{pmatrix} \bar{Y}_i \\ \bar{X}_i \end{pmatrix} : \sum_{i \in I} \lambda_i = 1, \lambda_i \geq 0 \right\}
 \end{aligned} \tag{6}$$

- Step 5. The least-norm projection to the observable frontier is the smallest of the D_i for each inefficient DMU.

3.3 Computational Complexity of the Proposed Method

The computational complexity of the least-norm projection method for n DMUs involves the solution of n linear programs. For the observable frontier projection, a non-linear program must be run for each unique hyperplane times the number of inefficient DMUs. As an illustration of the computational complexity of this method, an example involving 14 bank branches in Section 4 was solved on a Macintosh 9500 in 3.2 minutes of CPU time. Thus, the increased computational complexity of the proposed method is not a limiting factor in applying this method.

4. An Empirical Illustration

In order to provide a numerical illustration of the least-norm projection and observable frontier techniques, the Sherman and Gold (1985) data set on 14 bank branches is used. Sherman and Gold utilized DEA to evaluate the technical efficiency of branch operations. Their analysis is extended as well as the subsequent analysis of Haag and Jaska (1995), which used the same data set.

The data set consists of three inputs and four outputs over 14 branches. The input and output categories are as follows (for a complete description of these categories see Sherman and Gold (1985)):

Input #1	Rent (thousands of dollars)
Input #2	Full time equivalent personnel
Input #3	Supplies (thousands of dollars)
Output #1	Loan applications, new passbook loans, life insurance sales
Output #2	New Accounts, closed accounts
Output #3	Travelers checks sold, bonds sold, bonds redeemed
Output #4	Deposits, withdrawals, checks sold, treasury checks issued, B% checks, loan payments, passbook loan payments, life insurance payments, mortgage payments

Appendix A contains a listing of the data and the results of various DEA analyses. The original data is shown in Table A-1, and Table A-2 presents the scaled data using the technique introduced by Haag and Jaska (1985). The scaling technique divides each data point by the mean of all fourteen banks for that dimension. The scaling is necessary before a least-norm projection method is used, as there is not yet a scale invariant method. The scaling before hand is the best approximation of a scale invariant method. Table A-3 depicts efficiency scores for the input-oriented, output-oriented, and least-norm projection methods under the assumption of constant returns to scale. Table A-8 shows these results under the assumption of variable returns to scale. As can be seen from these tables, eight of the fourteen banks are efficient under the assumption of constant returns to scale while twelve of fourteen are efficient under the assumption of variable returns to scale.

To illustrate the vastly different implications that can come from the various DEA-based models, one need only compare the methods of Sherman and Gold (1985), Haag and Jaska (1995), as well as those developed in this paper. Sherman and Gold used the input-oriented, constant returns to scale DEA model to determine the efficiency of 14 bank branches. For Branch #7, Sherman and Gold determined that the DMU efficiency score was 0.782 and when compared to its reference set, and specific input decreases of 22%, 22% and 35% were required to the three inputs (while holding outputs constant) to make the branch efficient. It should be noted that Sherman and Gold did not use the efficiency score in their analysis but rather, used the reference set and the reference set weights to determine the reductions. From Table A-4, one can see that the reference set for Branch #7 includes Branches 3 and 6 at weights of 8% and 57% respectively. Because this is a constant returns to scale model, the origin is implicitly in every reference set, in this case at a weight of 35% ($=100-8-57$). Thus, Sherman and Gold deduce that the projection onto the frontier is not 22% ($=1-0.78$) of each input as would be the effect of reducing each input by the efficiency score (and used in most DEA analysis) but rather 22%, 22% and 35% as shown in Table 4.

Table 4. Sherman & Gold Method for Branch #7 Under Constant Returns to Scale

Inputs	Branch 3 (8%)	Branch 6 (57%)	Composite	Decrease
Rent	36,600	50,800	31,884	22%
FTE	14,200	8,300	5,867	22%
Supplies	29,800	18,900	13,157	35%

Haag and Jaska (1995) used a variable returns to scale method based on the multiplier dual to the additive DEA model to show that decreases of 3% to each input and increases of 9%, 5%, 5% and 3% to each output would make the branch efficient. These are obviously vastly different conclusions and it is important to understand the cause of these differences. Some of the difference can be explained by the returns to scale assumption. Using Sherman and Gold's (1985) methodology under a variable returns to scale assumption yields a DMU efficiency score of 0.93 and a reduction of inputs of 7% each to make this branch efficient. This is much closer to Haag and Jaska's recommendation, but is still significantly different due to the ability of the Haag and Jaska method to simultaneously move inputs and outputs. Unless there is an explicit restriction to hold outputs or inputs constant, the ability to adjust both simultaneously is superior in terms of determining what a branch would look like if efficient.

However, there are still problems with the Haag and Jaska method. Their method uses multiplier dual to the additive DEA model which builds the efficient frontier one facet at a time and in doing so, determines the least-norm projection from a given DMU to its associated facet. The problem is that although it accurately determines the shortest projection onto the current facet of the frontier, there is no assurance that this is the shortest projection to the entire frontier. In fact, in the case of the Sherman and Gold data, this projection is not the closest projection. If the shortest projection to the entire frontier is computed using the methodology described herein, then Branch #7 would need reductions of the inputs of 1.4%, 0.6%, and 17.1% while simultaneously increasing the outputs by 0.3%, 0.1%, 8.0%, and 0.1%. A summary of the results of the three models as well as the input-oriented DEA is shown in Table 5.

Table 5. Comparison of DMU Inefficiency by Model Type (VRS)*

	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4
Input-Oriented DEA	-7.0%	-7.0%	-7.0%	N/A	N/A	N/A	N/A
Sherman & Gold	-7.0%	-7.0%	-7.0%	N/A	N/A	N/A	N/A
Haag & Jaska	-3.0%	-3.0%	-3.0%	+9.0%	+5.0%	+5.0%	+3.0%
<i>least-norm projection</i>	-1.4%	-0.6%	-17.1%	+0.3%	+0.1%	+8.0%	+0.1%

* All models compare the inefficiency of Branch #7

This example shows that the choice of DEA model and the implication of that choice are very

serious in terms of managerial implications. Rather than requiring a slight decrease (3%) in rent, FTEs, and supplies and simultaneously requiring outputs to increase between 3% and 9%, the shortest projection onto the frontier shows that by concentrating on reducing supplies significantly, efficiency can be achieved without altering the other inputs and moderately altering Output #3. See Table A-11 for the percent change of each DMU when projecting onto the efficient frontier constructed under the assumption of variable returns to scale. The different recommendations that arise here highlight the importance of the appropriate reference point for an inefficient DMU. If the intent of a reference set is to compare against those DMUs on the frontier that most closely resemble the inefficient DMU, then, without compelling reasons to restrict any of the dimensions, the reference point should be the shortest projection.

As just described, there can be significantly different managerial implications based on the DEA method used when analyzing the efficiency of DMUs and thus, it is very important to choose the correct method of analysis. Again, we believe that if the situation explicitly restricts either inputs or outputs from varying, then the classic oriented methods are recommended. However, if the situation does not restrict the reduction of inputs or the augmentation of outputs, then the least-norm projection method presented herein is preferred. The proposed method is superior to Haag and Jaska's in that their method looks for the least-norm projection to a particular facet of the frontier, while our method takes the least-norm projection to the overall frontier. As shown above, this small difference in methodology can lead to drastically different results.

The second point to be made in this section deals with the projection onto the *observable* frontier. Table 4 depicts the original data, the scaled data, the shortest projection to the overall frontier, and the scaled data for each of the 5 branches in Branch #7's reference set. From Table 3, one can see that the projection to the overall frontier requires a significant decrease in supplies (Input #3), a moderate increase in Output #3, and virtually no change along the other five dimensions. However, from Table 4, one sees that the recommended reduction in supplies is to a level not experienced by any of the branches in the reference set. In fact, the recommendation is to a level (0.5892) beneath which any of the fourteen branches in the sample have experienced. While this may be a plausible level of reduction, it is likely that another useful set of information would be the closest projection on the frontier restricted to what has been experienced by other branches or, in other words, within the convex hull of what has been experienced by other branches. Thus, it is useful to add to this analysis the shortest projection to the observable frontier (see Section 3 for a full description of the observable frontier) as shown in Table 5. As can be noted from this table, all of the projections are within the range of the reference set's experience. Recall that the reference set comprises all DMUs that lie on the facet where the shortest projection lands. Thus, it is quite likely, and in fact true in this case, that the reference set changes when projecting onto the observable portion of the frontier. When looking at a comparison of the two projections, as shown in Table 6, it is clear that the recommendations stemming from a projection to the overall frontier can be significantly different than the recommendations stemming from a projection onto the observable frontier.

Table 6. Branch #7 Reference Set Information for Shortest Projection (VRS)

	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
Original Data	40,800	7,500	20,400	53,000	465,700	20,240	1,800,250
Scaled Data	0.8228	0.6406	0.7104	0.2635	0.5040	0.4424	0.7623
Projection	0.8221	0.5312	0.7097	0.2642	0.5734	0.4431	0.7864
<i>Reference Set:</i>							
2	0.9842	1.4863	1.3199	1.9091	1.8242	3.0551	1.4130
6	1.0245	0.7090	0.6582	1.3523	0.7223	0.7595	1.1264
8	0.6331	0.7632	0.7436	1.2446	0.6973	0.9476	0.9928
11	0.8971	0.7163	0.7554	0.5223	0.5224	1.0854	0.9624
13	0.8183	0.4269	0.6751	0.4181	0.3116	0.7757	0.5023

Table 7. Branch #7 Reference Set Information for Shortest Observable Projection (VRS)

	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
Original Data	40,800	7,500	20,400	53,000	465,700	20,240	1,800,250
Scaled Data	0.8221	0.5312	0.7097	0.2642	0.5734	0.4431	0.7864
Projection	0.7298	0.4553	0.6778	0.4504	0.4647	0.4818	0.5856
<i>Reference Set:</i>							
5	0.6572	0.3630	0.6824	0.2289	0.4204	0.4137	0.4593
6	1.0245	0.7090	0.6582	1.3523	0.7223	0.7595	1.1264
9	0.7341	0.6492	0.7313	2.0234	0.701	0.7073	0.8393

Table 8. Comparison of Least-Norm Projections (VRS)

	Inputs			Outputs			
	Rent	FTEs	Supplies	Output #1	Output #2	Output #3	Output #4
Projection to Overall Frontier:	-1.4%	-0.6%	-17.1%	+0.3%	+0.1%	+8.0%	+0.1%
Projection to Observable Frontier:	-11.3%	-28.9%	-4.6%	+70.9%	+7.8%	+8.9%	+23.2%

5. Contribution & Future Research

This paper has focused on the importance of selecting the appropriate benchmarking or reference set when making recommendations for improving inefficient DMUs. As has been shown in the preceding sections, different projections onto the efficient frontier yield different reference sets and thus, provide for varying managerial recommendations. The method described herein improves upon existing work as it allows an inefficient DMU to benchmark against those efficient DMUs that are more similar to itself in terms of input and output consumption. That is, by taking the shortest projection to the efficient frontier, as opposed to a restricted projection, an inefficient DMU is guaranteed to benchmark against those DMUs most similar to itself.

There have been two major contributions made by this work. Each of these contributions comes as an extension of the basic DEA methodology in order to reflect more meaningful projections onto the efficient frontier. The first extension is to offer a new methodology that uses the least-norm projection to the efficient frontier. This projection is more meaningful than either the input- or output-oriented projections as it permits the simultaneous movement of inputs and outputs. The second contribution is in developing a methodology to project onto the observable portion of the efficient frontier. That is, when taking the shortest projection, it is often the case that the projection lands beyond the experience of any existing DMUs. Thus, it is useful to project onto the portion of the frontier that has either been experienced by existing DMUs or is a convex combination of existing DMUs.

Although the methodology developed herein is very useful, there is a shortcoming that is left to future research; namely, that the proposed method is not invariant with respect to the scale of the units used for the inputs and/or outputs. That is, in the absence of applying a scaling technique ahead of time, the results of this method vary as the units of the inputs and outputs vary. In fact, any projection-based method will suffer from this lack of units invariance. This paper used the scaling technique proposed by Haag and Jaska (1995) to overcome this shortcoming, but it remains an open question to develop units-invariant projection methods.

Acknowledgments

This work was made possible by a grant from the Alfred P. Sloan Foundation to the Wharton Financial Institutions Center.

References

- Ali, A.I. and Seiford, L.M. (1993), "The mathematical programming approach to efficiency analysis." In *The Measurement of Productive Efficiency*. Harold O. Fried, C. A. Knox Lovell and Shelton S. Schmidt (eds). New York: Oxford University Press.
- Banker, R. and Morey, R.C. (1986), "The use of categorical variables in data envelopment analysis," *Management Science*, 32/12, 1613-1627.
- Charnes, A. and Cooper, W. (1980), "Management science relations for evaluation and management accountability", *Journal of Enterprise Management*, 2, 143-162.
- Charnes, A., Cooper, W.W., Golany, B., Seiford, L., and Stutz, J. (1985), "Foundations of data envelopment analysis for pareto-koopmans efficient empirical production functions", *Journal of Econometrics*, 30/1, 91-107.
- Charnes, A., Cooper, W. and Rhodes, E. (1978), "Measuring efficiency of decision-making units", *European Journal of Operational Research*, 2/6, 428-449.
- Charnes, A., Cooper, W. and Rhodes, E. (1982), "A multiplicative model for efficiency analysis", *Socio-Economic Planning Sciences*, 16/5, 223-224.
- Charnes, A., Haag, S., Jaska, P., and Semple, J. (1992), "Sensitivity of efficiency classifications in the additive model of DEA", *International Journal of Systems Science*, 23/5, 789-798.
- Charnes, A., Rousseau, J., and Semple, J. (1995), "Sensitivity of classifications in the ratio model of DEA", *Journal of Productivity Analysis*, 7/1, 5-18.
- Dyson, R.G. and Thanassoulis, E. (1988), "Reducing weight flexibility in data envelopment analysis", *Journal of the Operational Research Society*, 39, 563-576.
- Fare, R., Grosskopf, S. and Lovell, C.A.K (1985), "The Measurement of efficiency of production." Boston: Kluwer-Nijhoff Publishing.
- Haag, S. and Jaska, P. (1995), "Interpreting inefficiency ratings: an application of bank branch operating efficiencies", *Managerial and Decision Economics*, 16, 7-14.
- Land, K.C., Knox Lovell, C.A., and Thore, S. (1993), "Chance-constrained data envelopment analysis", *Managerial and Decision Economics*, 14, 541-554.
- Ortega, J.M., and Rheinbaldt, W.C. (1970), "Iterative Solution of Nonlinear Equations in Several Variables." New York: Academic Press.
- Rousseau, J., and Semple, J. (1995), "Radii of classification preservation in DEA: A case study of program follow through", *Journal of the Operational Research Society*, 46/8, 943-957.

Rousseau, J.J. and Semple, J.H. (1993), "Notes: Categorical outputs in data envelopment analysis," *Management Science*, 39/3, 384-386.

Seiford, L. (1993), "A bibliography of Data Envelopment Analysis (1978-1993)", Technical Report, Department of Industrial Engineering , University of Massachusetts, Amherst.

Sengupta, Jati K. (1989), "Nonlinear measures of technical efficiency", *Computers & Operations Research*, 16/1, 55-65.

Sherman, H. and Gold, F. (1985), "Bank branch operating efficiency", *Journal of Banking and Finance*, 9, 297-315.

Sueyoshi, Toshiyuki (1994), "Stochastic frontier production analysis: Measuring performance of public telecommunications in 24 OECD countries", *European Journal of Operational Research*, 74/3, 466-478.

Appendix A. Tables for Sherman & Gold (1985) Data

Table A-1 depicts the input and output data originally used by Sherman and Gold (1985) in evaluating the operating performance of fourteen bank branches.

Table A-1 Original Data

Branch	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
1	140,000	42,900	87,500	484,000	4,139,100	59,860	2,951,430
2	48,800	17,400	37,900	384,000	1,685,500	139,780	3,336,860
3	36,600	14,200	29,800	209,000	1,058,900	65,720	3,570,050
4	47,100	9,300	26,800	157,000	879,400	27,340	2,081,350
5	32,600	4,600	19,600	46,000	370,900	18,920	1,069,100
6	50,800	8,300	18,900	272,000	667,400	34,750	2,660,040
7	40,800	7,500	20,400	53,000	465,700	20,240	1,800,250
8	31,900	9,200	21,400	250,000	642,700	43,280	2,296,740
9	36,400	7,600	21,000	407,000	647,700	32,360	1,981,930
10	25,700	7,900	19,000	72,000	402,500	19,930	2,284,910
11	44,500	8,700	21,700	105,000	482,400	49,320	2,245,160
12	42,300	8,900	25,800	94,000	511,000	26,950	2,303,000
13	40,600	5,500	19,400	84,000	287,400	34,940	1,141,750
14	76,100	11,900	32,800	199,000	694,600	67,160	3,338,390

Table A-2 depicts the scaled branch data which divides each data point by the average of the fourteen branches on each dimension. This technique was suggested by Haag and Jaska (1995) in order to compensate for the fact that the least-norm projection techniques are not units invariant. By using this scaling technique before running DEA, the data is units invariant.

Table A-2 Scaled Data

Branch	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
1	2.8234	3.6644	3.0473	2.4063	4.4798	1.3083	1.2498
2	0.9842	1.4863	1.3199	1.9091	1.8242	3.0551	1.4130
3	0.7381	1.2129	1.0378	1.0391	1.1461	1.4364	1.5118
4	0.9499	0.7944	0.9333	0.7805	0.9518	0.5975	0.8814
5	0.6572	0.3630	0.6824	0.2289	0.4204	0.4137	0.4593
6	1.0245	0.7090	0.6582	1.3523	0.7223	0.7595	1.1264
7	0.8221	0.5312	0.7097	0.2642	0.5734	0.4431	0.7864
8	0.6331	0.7632	0.7436	1.2446	0.6973	0.9476	0.9928
9	0.7341	0.6492	0.7313	2.0234	0.701	0.7073	0.8393
10	0.5183	0.6748	0.6617	0.3580	0.4356	0.4356	0.9676
11	0.8971	0.7163	0.7554	0.5223	0.5224	1.0854	0.9624
12	0.8515	0.6200	0.8969	0.4689	0.5547	0.6281	1.0365
13	0.8183	0.4269	0.6751	0.4181	0.3116	0.7757	0.5023
14	1.5347	1.0165	1.1423	0.9893	0.7518	1.4679	1.4137

Table A-3 depicts the efficiency scores for each of the oriented DEA methods as well as the least-norm projection method under the assumption of constant returns to scale.

Table A-3 Efficiency Scores for Constant Returns to Scale

Branch	Input-Oriented	Output-Oriented	Least-Norm Projection
1	1.00	1.00	0.00
2	1.00	1.00	0.00
3	1.00	1.00	0.00
4	1.00	1.00	0.00
5	0.90	1.11	0.04
6	1.00	1.00	0.00
7	0.78	1.28	0.13
8	0.98	1.02	0.03
9	1.00	1.00	0.00
10	1.00	1.00	0.00
11	0.97	1.03	0.03
12	0.85	1.17	0.15
13	0.90	1.11	0.05
14	1.00	1.00	0.00

Table A-4 depicts the reference set for each DMU using input-oriented DEA and output-oriented DEA under the assumption of constant returns to scale.

Table A-4 Oriented Reference Sets (CRS)

Branch	Input-Oriented Reference Set (CRS)				Output-Oriented Reference Set (CRS)			
1	1				1			
2	2				2			
3	3				3			
4	4				4			
5	2 (0.05)	4 (0.17)	6 (0.21)		2 (0.06)	4 (0.18)	6 (0.23)	
6	6				6			
7	3 (0.08)	6 (0.57)			3 (0.10)	6 (0.73)		
8	2 (0.04)	3 (0.36)	6 (0.07)	9 (0.36)	2 (0.04)	3 (0.36)	6 (0.08)	9 (0.34)
9	9				9			
10	10				10			
11	2 (0.16)	6 (0.39)	14 (0.20)		2 (0.16)	6 (0.40)	14 (0.21)	
12	3 (0.03)	6 (0.49)	10 (0.38)		3 (0.04)	6 (0.58)	10 (0.45)	
13	2 (0.16)	14 (0.18)			2 (0.18)	14 (0.20)		
14	14				14			

Table A-5 depicts the reference set for each DMU using the least-norm projection and the shortest *observable* distance under the assumption of constant returns to scale. The shortest observable distance is the least-norm projection to the observable portion of the frontier. The reference set in this case is the set of DMUs that are on the facet of the frontier that the projection lands on.

Table A-5 Least-Norm and Shortest Observable Distance Reference Set (CRS)

Branch	Least-Norm Projection Reference Set (CRS)			Shortest Observable Distance Reference Set (CRS)		
1	1	2		1	2	
2	2	9		2	9	
3	2	3	9	2	3	9
4	2	4	6	2	4	6
5	2	4	6	3	6	10
6	6	9		6	9	
7	2	4	6	3	6	10
8	3	6	10	3	6	10
9	9			9		
10	3	6	10	3	6	10
11	2	6	14	3	6	10
12	2	6	14	3	6	10
13	2	6	14	3	6	10
14	2	6	14	2	6	14

Table A-6 depicts the percent change from each DMU to its projection onto the efficient frontier, under the assumption of constant returns to scale. Obviously, a percent change of zero across all dimensions implies that the DMU is efficient.

Table A-6 Shortest Projection Change (CRS)

Branch	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	-7.6%	0.0%	0.1%	4.7%	0.0%	1.5%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	-0.1%	-17.1%	-0.1%	0.3%	13.8%	0.2%	3.2%
8	-1.6%	-2.9%	-0.2%	0.1%	0.2%	0.2%	2.1%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
11	0.0%	-3.6%	0.0%	0.1%	0.1%	0.7%	1.2%
12	-0.2%	-18.4%	-0.2%	0.3%	0.3%	6.6%	6.3%
13	-0.1%	-9.1%	-0.1%	0.1%	0.2%	1.6%	3.9%
14	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A-7 depicts the percent change from each DMU to its projection onto the observable portion of the efficient frontier, under the assumption of constant returns to scale. Obviously, a percent change of zero across all dimensions implies that the DMU is efficient.

Table A-7 Shortest Observable Projection Change (CRS)

Branch	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	-21.2%	71.7%	-3.1%	56.5%	8.5%	5.3%	113.7%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	-35.0%	5.5%	-6.9%	48.3%	-11.7%	0.9%	27.6%
8	36.4%	1.8%	-1.8%	-10.5%	8.2%	-12.1%	20.5%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
11	-28.2%	12.1%	1.8%	29.5%	29.2%	-29.3%	20.2%
12	-28.4%	-1.9%	-21.2%	22.2%	1.4%	2.1%	8.2%
13	-28.3%	44.6%	-2.1%	18.1%	52.5%	-37.2%	104.6%
14	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A-8 depicts the efficiency scores for each of the oriented DEA methods as well as the least-norm projection method, under the assumption of variable returns to scale.

Table A-8 Efficiency Scores for Variable Returns to Scale

Branch	Input-Oriented	Output-Oriented	Least-Norm Projection
1	1.00	1.00	0.00
2	1.00	1.00	0.00
3	1.00	1.00	0.00
4	1.00	1.00	0.00
5	1.00	1.00	0.00
6	1.00	1.00	0.00
7	0.94	1.23	0.12
8	1.00	1.00	0.00
9	1.00	1.00	0.00
10	1.00	1.00	0.00
11	1.00	1.00	0.00
12	0.87	1.16	0.20
13	1.00	1.00	0.00
14	1.00	1.00	0.00

Table A-9 depicts the reference set for each DMU using input-oriented and output-oriented DEA under the assumption of variable returns to scale.

Table A-9 Oriented Reference Sets (VRS)

Branch	Input-Oriented Reference Set (VRS)					Output-Oriented Reference Set (VRS)				
1	1					1				
2	2					2				
3	3					3				
4	4					4				
5	5					5				
6	6					6				
7	5 (0.16)	6 (0.34)	10 (0.30)	13 (0.19)		5 (0.16)	6 (0.47)	9 (0.20)	10 (0.17)	
8	8					8				
9	9					9				
10	10					10				
11	11					11				
12	5 (0.11)	6 (0.39)	8(0.03)	10 (0.45)	11 (0.02)	3 (0.12)	6 (0.61)	10 (0.27)		
13	13					13				
14	14					14				

Table A-10 depicts the reference set for each DMU using the least-norm and shortest observable distance under the assumption of variable returns to scale. The shortest observable distance is the least-norm projection to the observable portion of the frontier. The reference set in this case is the set of DMUs that are on the facet of the frontier that contains the projection.

Table A-10 Least-Norm and Shortest Observable Distance Reference Set (VRS)

Branch	Least-Norm Projection Reference Set					Shortest Observable Distance Reference Set				
1	1	2				1	2			
2	2					2				
3	2	3	8	9		2	3	8	9	
4	2	4	9			2	4	9		
5	5	9	13			5	9	13		
6	6	9				6	9			
7	2	6	8	11	13	5	6	9		
8	2	8	9			2	8	9		
9	2	9				2	9			
10	8	9	10			8	9	10		
11	2	6	8	11	13	2	6	8	11	13
12	3	6	14			8	9	10		
13	5	9	13			5	9	13		
14	3	6	14			3	6	14		

Table A-11 depicts the percent change from each DMU to its projection onto the efficient frontier, under the assumption of variable returns to scale. Obviously, a percent change of zero across all dimensions implies that the DMU is efficient.

Table A-11 Shortest Projection Change (VRS)

	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	-1.4%	-0.6%	-17.1%	0.3%	0.1%	8.0%	0.1%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
11	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
12	-0.9%	-16.6%	-0.8%	1.6%	1.4%	1.3%	15.5%
13	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
14	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A-12 depicts the percent change from each DMU to its projection onto the observable portion of the efficient frontier, under the assumption of variable returns to scale. Obviously, a percent change of zero across all dimensions implies that the DMU is efficient.

Table A-12 Shortest Observable Projection Change (VRS)

	Inputs			Outputs			
	Rent	FTEs	Supplies	Out 1	Out 2	Out 3	Out 4
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	-11.3%	-28.9%	-4.6%	70.9%	-7.8%	8.9%	-23.2%
8	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
11	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
12	-35.7%	-7.7%	-24.1%	22.3%	-9.9%	-5.1%	-0.7%
13	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
14	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%