USING DATA ENVELOPMENT ANALYSIS TO SELECT EFFICIENT LARGE MARKET CAP SECURITIES

Jennifer Powers, University of Maine Patrick R. McMullen, Auburn University

This research presents a technique known as Data Envelopment Analysis (DEA) to select a group of desirable securities from a list of 185. This selection process is an application of Operations Research, and reflects the utility function of a decision-maker. This utility function considers many attributes associated with large cap stocks such as return, earnings and risk. Of the 185 stocks analyzed, fourteen were found to be most desirable (referred to here as efficient), and an additional four were found to be somewhat desirable. Of these fourteen "efficient" stocks, several displayed robustness to unfavorable changes, while others did not.

The decision to purchase stock can be difficult since there are many attributes to consider. Some of these attributes may include (but are not limited to) return on investment, earnings per share, price to earnings ratio, and risk. While some investors might be more concerned with minimizing the level of risk in their portfolio, others may be more concerned with a high return on their investment. Since the decision to purchase stock can include the necessary examination of several attributes, it can be thought of as a multi-criteria decision-making problem.

Data Envelopment Analysis (DEA) can assist with such problems as described above. DEA is a multi-criteria decision-making technique that can select the "most favorable," or desirable alternatives from a large set when it is necessary to consider several attributes, given, of course, the utility function of the decision-maker. DEA also helps to minimize the complexity of analysis by simultaneously evaluating the attributes of interest and presenting a single, composite score, referred to as efficiency. Efficiency, as used in this paper, describes the alternative(s) with a set of attributes that collectively dominate the others based on the simultaneous analysis of all alternatives and their attributes as described above. The efficiency value of an alternative is the objective function value of a mathematical programming model. An alternative is deemed DEA-efficient if its costs (inputs) are offset by its benefits (outputs). Otherwise, an alternative is classified as DEA-inefficient.

DEA is a tool that has gained popularity over the last several years. It can be used as a tool assisting with classification between efficient and inefficient alternatives for a variety of applications, where data takes on either a subjective or empirical form. Charnes, Cooper, Lewin and Seiford (1994) present several of these types of applications in their DEA book.

This research examines 185 of the largest market cap stocks. A total of eight attributes are considered, five of the attributes are considered outputs: 1, 3, 5, & 10 year return and earnings per share, while the remaining three attributes are considered inputs: price to

earnings ratio, beta, and sigma. In the context of this paper, "outputs" are assumed to be benefits associated with owning a specific stock, and "inputs" are assumed to be costs associated with owning a specific stock. The reason that inputs and output are considered here is because the most general definition of "efficiency" typically considers the relationship between inputs and outputs. The stocks that are found to be "best" (efficient) are examined further for robustness to unfavorable changes in their attributes. Several of the stocks that were labeled as inefficient (but still have somewhat desirable attributes) are examined to determine how much favorable change in their attributes is needed in order for them to be considered efficient.

It should be noted that this study does not claim to necessarily be a "stand alone" decision-making tool, but instead a source of information showing how Data Envelopment Analysis can be used to discriminate between strong performers and others. It is also intended to provide information as to how close a stock classified as DEA-inefficient actually is to being classified as DEA-efficient.

DATA AND ATTRIBUTES

This analysis evaluates 185 of the largest market cap stocks. The market cap for a stock is defined as the total amount of money invested in outstanding shares of stock. It can be calculated by taking the number of a firm's outstanding stocks and multiplying it by the selling price of that stock. For instance, if a firm has 1,000,000 shares of stock outstanding and their stock sells for \$60/share, then their market cap value is \$60,000,000 (1,000,000 * 60).

As mentioned, there are eight attributes used for this analysis. The data was collected using the Value-Line Investment Survey (Tobel, 1998). Stocks with missing data were eliminated from this analysis if follow-up efforts were unsuccessful in finding some of this missing data. Web-sites were used as a follow-up to find missing data.

The eight attributes considered, their input or output classifications, and general definitions are found in Table 1.

Attribute	Classification	Definition
1-Year Return	Output	Average return over a 1 year period in %
3-Year Return	Output	Average return over a 3 year period in %
5-Year Return	Output	Average return over a 5 year period in %
10-Year Return	Output	Average return over a 10 year period in %
Earnings per share	Output	Ratio of earnings to the number of shares held
Price to earnings ratio	Input	Stock price divided by earnings per share
Beta (risk) (5 year)	Input	Relationship to rest of market
Sigma (5 year)	Input	3-year standard deviation of returns

Table 1
Description of Attributes

The classification column provides a general classification of all eight attributes as either an input or an output. The outputs, 1, 3, 5, & 10-year return and earnings per share

are considered such because they are benefits gained from owning a certain stock. The investor would want high values for these attributes. Price to earnings ratio, beta and sigma are all considered inputs because they are something the investor must give up in order to own that stock, i.e. high prices, high levels of risk, etc. The investor would want the values of these attributes to be as low as possible. Of course, the authors concede the fact that one could make reasonable arguments for classifying these outputs as inputs, etc., due to such real-world concerns as taxation, etc. (after all, higher returns suggest higher taxation, etc.). Nonetheless, in deciding which attributes should be classified as inputs or outputs, the authors decided in favor of treating attributes with benefits as outputs, and treating attributes with costs as inputs.

METHODOLOGY

Data Preparation

The eight attributes presented above are in differing units (for example, EPS is a ratio, while several other measures are percentages). Although Data Envelopment can successfully process attributes on differing numeric scales via weighting, standardization is performed on attributes here so that weighting schemes can be kept as universal as possible. Composite interpretation of these attributes in their present form can only be properly interpreted if pairwise weighting schemes are carefully constructed so that overly-binding constraints can be avoided. Prior to analysis, the attribute values for each security must be converted into a standardized form. To explain the standardization process, the following definitions are provided:

 X_{ij} = the value of attribute j for stock i

 X_j = the mean attribute value for attribute j across all i stocks

 $\hat{\sigma}_j$ = the standard deviation of the value for attribute j

Zij = the number of standard deviations stock i is above or below the mean for attribute j (Z-scores).

RZij = the Z-scores re-scaled from zero.

Determination of the mean and standard deviation across all stocks for each attribute is straightforward. The values of Z_{ij} (the Z-scores) are determined by the following formula:

$$Z_{ij} = \frac{(X_{ij} - \overline{X}_j)}{\hat{\sigma}_j} \tag{1}$$

After these Z-scores are determined, they need to be re-scaled from zero for each attribute, so that the minimum Z-score for each attribute is zero. It is important to have the minimum Z-scores for each attribute be zero because the objective function of the DEA

model is constructed with the requirement of minimum inputs and outputs having values of zero. This is done by adding the absolute value of the minimum Z-score to each Z-score for each attribute:

$$RZ_{ij} = Abs(Min Z_j) + Z_{ij}$$
, for each attribute j (2)

After these data are standardized and re-scaled from zero, they are ready for analysis.

Data Envelopment Analysis Formulation

Data Envelopment Analysis is employed to determine which securities are efficient. A security is considered efficient if its composite measure of output offsets its corresponding composite measure of input (Banker et al, 1984 and Charnes et al, 1978 and 1994). Recall that for this analysis, the attributes of EPS, and the returns (1, 3, 5, and 10 year) are considered outputs, while the PE ratio, Beta and Sigma measures are considered inputs. Table 2 below shows the variable names for each specific output and input in terms of both their values and weights:

Attribute Value Symbol Weight Symbol **EPS** Ο, 1 Year Return Ο, 3-Year Return O_3 u_3 5-Year Return O_4 u_4 10-Year Return PE Ratio Beta V 2 Sigma

Table 2
Variables for Inputs, Outputs and Weights

In Data Envelopment Analysis, the objective is to maximize the composite efficiency of each decision-making unit — the efficiency of decision-making unit i is denoted as h_i. In this case, the decision-making units are stocks. This objective function uses linear programming to find the set of weights that maximize this composite function. Mathematically, this is as follows:

Max:
$$h_i = \sum_{i=1}^{5} u_{ij} O_{ij} + \mu_i$$
, for each stock i (3)

Subject to the following constraints:

$$\sum_{j=1}^{3} v_{ij} I_{ij} + \rho_i = 1, \text{ for each stock i}$$
(4)

and:

$$\sum_{j=1}^{5} u_{ij} O_{ij} - \sum_{j=1}^{3} v_{ij} I_{ij} + \mu_{i} \le 0, \text{ for each stock i}$$
 (5)

The values of μ_i and ρ_i are unrestricted slack variables which quantify the amount of inefficiency present in each stock i. All weights are restricted to be at least zero, and upper bound restrictions are discussed in the next section. This analysis is performed for each of the 185 stocks.

The objective function is intended to provide the maximum efficiency value for stock i, given the weighting factors of each output. Alternatives found to be DEA-efficient will have values of 100% according to equation (3). The two constraints (equations (4) and (5)) are intended to work in tandem to ensure that the efficiency of any security never exceeds unity.

In the most general terms, the objective is to show each security in its best possible light (maximization of h_i), while simultaneously adhering to the rules that restrict a security from having an efficiency exceed unity (equations (4) and (5)).

Weighting Issues

The issue of weighting needs additional consideration. Consider the values of the weights. The values of these weights must be greater than or equal to zero, but this restriction by itself is not adequate. Consider the following example: a security has desirable returns and a strong EPS ratio, but it is volatile — values of Beta and Sigma are high. Obviously, the high level of volatility detracts from the relative desirability of this particular security. This security, however, would most likely be shown as favorable according to the above linear programming model simply because the weights for Beta and Sigma (v₂ and v₃) would be permitted to take on a weight of zero, thereby essentially ignoring the relative importance of these attributes. This condition does not allow a reliable gauge of efficiency for the securities considered.

To address this concern, additional constraints need to be incorporated so that undesirable levels of certain attributes are not completely ignored (Thompson et al, 1990 and Wong et al, 1990). Constructing pairwise relationships of weights for the general output attributes does address this condition. These additional weighting constraints are as follows:

$$.2 \le \frac{u_1}{u_2}, \frac{u_1}{u_3}, \frac{u_1}{u_4}, \frac{u_1}{u_5}, \frac{u_2}{u_5}, \frac{u_2}{u_4}, \frac{u_2}{u_5}, \frac{u_3}{u_4}, \frac{u_3}{u_5}, \frac{u_4}{u_5} \le 5$$
 (6)

Similarly, the input weighting restrictions are as follows:

$$.2 \le \frac{v_1}{v_2}, \frac{v_1}{v_3}, \frac{v_2}{v_3}, \le 5 \tag{7}$$

These constraints (equations (6) and (7)) essentially state that any attribute may be considered no more than five times as important as other attributes with respect to the individual sets of outputs and inputs. This weighting restriction factor of five is admittedly

general, but is intended to have all attributes represented, while at the same time, permit the stronger attributes to have greater representation. The concluding section of this paper addresses this thought in the context of weighting with respect to the decision-maker's utility function.

The fact that these pairwise weighting restrictions are generalized as such can be attributed to the standardization and subsequent re-scaling of the model data. In other words, because these data has been standardized to the same scale, there is an ability to declare generalized pairwise weighting restrictions. This would not have been possible if these data were not standardized.

It is also appropriate to note that the four return attributes (1, 3, 5, and 10 year) are correlated with each other. The presence of this correlation is not a source of concern here as it might be in the case of multicollinearity for a multiple regression analysis. The reason for this is because in this analysis, the investor is extracting utility from these returns, and the DEA model is simply finding the best combination of returns (subject to the weights) so that the efficiency score is maximized. Interpretation of model results is not threatened here as it might be in a multiple regression analysis (McMullen and Strong, 1998).

RESULTS

Efficiency

Below is a list of the stocks that were found to be DEA-efficient within the context of the previously described model — these stocks all have an efficiency of 100% as according to equation (3). Also listed with each stock are their input and output attributes.

Stock	Ticker	P/E	Beta	Std	EPS	1Yr Ret	3Yr Ret	5Yr Ret	10yr Ret
Exxon	XON	21.1	0.9	13.6	3.0	9.7	27.9	19.2	15.7
British Telecom	BTY	26.8	0.7	22.5	5.0	104.1	39.8	23.5	19.7
Dell Computer	DELL	65.6	1.3	53.9	1.7	143.7	174.9	144.3	76.9
Mobil	MOB	19.4	0.8	14.1	3.7	-2.0	17.0	16.3	17.1
AEGON Ins. Group	AEG	30.4	0.9	18.7	3.0	133.4	77.9	59.5	41.8
Vodafone Group ADR	VOD	61.6	1.1	28.1	2.0	149.6	47.4	37.1	108.3
Wells Fargo	WFC	21.8	1.2	22.5	14.0	12.7	17.0	22.0	20.2
Duke Energy	DUK	20.9	0.7	14.4	2.9	34.1	20.4	13.0	16.7
Safeway	SWY	29.4	0.8	25.6	1.4	52.9	58.2	52.6	176.0
America Online	AOL	164.5	1.6	55.9	0.5	162.0	72.5	96.2	148.7
Southern Company	SO	15.5	0.7	16.5	1.7	40.8	16.9	11.2	17.6
Canon, Inc.	CANNY	20.6	0.6	22.0	1.0	-26.0	5.0	9.3	8.7
Gen'l Re Corp.	GRN	16.4	1.0	20.2	13.0	8.1	13.1	11.0	16.2

Table 3
DEA-Efficient Stocks

Basically, these stocks have a collective set of attributes that essentially dominate the others. In addition to these stocks found to be DEA-efficient, there are a few others that,

20.6

1.7

44.9

8.7

3.2

13.2

0.6

18.8

PCG

PG&E Corp.

while not found to be DEA-efficient (referred to as DEA-inefficient), nonetheless have a collective set of attributes which are close to being DEA-efficient. This list is as follows:

Stock	Ticker	P/E	Beta	Std	EPS	1Yr Ret	3Yr Ret	5Yr Ret	10yr Ret
British Petroleum	BP	17.56	0.75	18.0	4.21	-11.06	21.48	24.33	16.39
Honda Motor ADR	HMC	16.16	0.65	26.3	4.20	11.31	25.11	19.97	9.09
Texaco	TX	20.30	0.70	16.0	2.82	-0.64	23.94	15.98	16.23
Texas Utilities	TXU	15.97	0.70	15.7	2.61	28.85	13.22	4.23	11.99

Table 4
Near-Efficient Stocks

These four securities are classified as "near-efficient," which means that they have a collective set of attributes that are not DEA-efficient, but nonetheless have attributes that are relatively desirable. It should be noted that selection of these four securities as being near-efficient was subjective — the authors found these four securities to have reasonably desirable levels of both inputs and outputs.

Table 5 below presents some information describing how much favorable change is needed to make these four securities DEA-efficient. The amount of required input reduction and output augmentation can also be thought of as the degree of inefficiency.

Table 5
Input Reduction, Output Augmentation for Near-Efficient Stocks

Stock	Ticker	Input Reduction	Output Augmentation
British Petroleum	BP	87.01%	11.86%
Honda Motor ADR	HMC	91.78%	6.57%
Texaco	TX	79.01%	19.34%
Texas Utilities	TXU	92.36%	16.37%

This idea of favorable change needs some explanation. The input "reduction column" shows how much the virtual inputs need to be reduced from their current levels to make the security of interest DEA efficient. In this context, virtual inputs is stated as follows:

Virtual Inputs =
$$\sum_{i=1}^{3} v_{ij} I_{ij}$$
 (8)

The "output augmentation" column shows how much the virtual outputs need to be augmented, or increased above their current levels so that the security of interest is considered DEA-efficient. Mathematically, virtual outputs is stated as follows:

Virtual Outputs =
$$\sum_{j=1}^{5} u_{ij} O_{ij}$$
 (9)

For these near-efficient stocks to be considered DEA-efficient, either the specified level of input reduction *or* output augmentation must occur, but both are not necessary. For example, for British Petroleum (BP) to be classified as DEA-efficient, their inputs must be reduced to 87.01% of their current levels *or* their outputs must be increased to 11.86% above their current levels.

Another way to interpret the input reduction and output augmentation requirements is to consider the alternative as excessive in terms of input and/or deficient in terms of output. For example, British Petroleum is excessive in terms of P/E ratio, Beta and standard deviation of return, while it is considered deficient in terms of EPS and returns. If input reduction is somehow performed to the DEA-inefficient alternative of interest to the specified level, or if output augmentation is performed to the alternative of interest (to the specified level), the alternative could then be considered efficient. In short, input reduction and output augmentation simply informs as to how much improvement is needed for the alternative of interest to be classified as DEA-efficient.

All of the securities found to be DEA-efficient reside on the "efficiency frontier." This efficiency frontier is a convex collection of points which classify DEA-efficient securities from ones that are DEA-inefficient — efficiency is 100%. DEA-efficient securities lie on the curve, while DEA-inefficient securities lie "southeast" of the frontier. Figure 1 shows the efficiency frontier for the securities found to be DEA-efficient, in addition to the near-efficient securities. Due to scaling reasons, Texas Utilities (TXU) is not shown as one of the near-efficient securities.

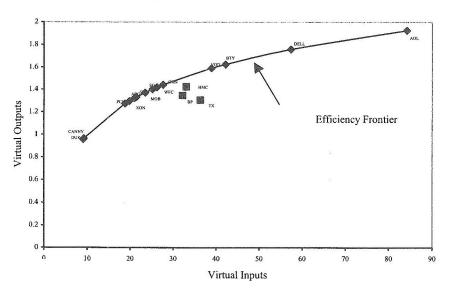


Figure 1. Efficiency Frontier

Sensitivity Analysis

In addition to determination of whether-or-not a security is found to be DEA-efficient, it is also important to examine the robustness of the DEA-efficient securities to unfavorable

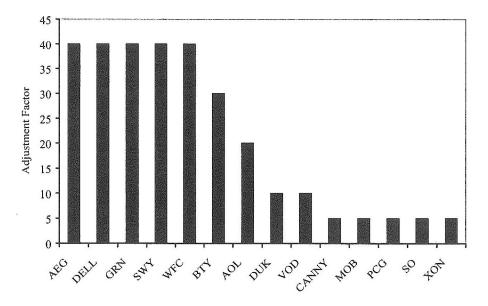
change. Specifically, it is desired to determine how long a DEA-efficient security will remain in the set of DEA-efficient securities as they are subjected to unfavorable change. In this context, unfavorable change means that for the DEA-efficient securities, the inputs are increased by some adjustment factor, and the outputs are decreased by some adjustment factor. Mathematically, this is as follows:

For inputs:
$$X_{ij} = X_{ij} * (1 + Adjustment Factor)$$
 (10)
For outputs: $X_{ij} = X_{ij} / (1 + Adjustment Factor)$ (11)

For outputs:
$$X_{ij} = X_{ij} / (1 + Adjustment Factor)$$
 (11)

These adjustments make inputs more expensive and outputs less beneficial. In short, the changes are unfavorable. Each time one of these changes is made to the set of DEAefficient securities, the analysis is repeated to see which of the securities initially found to be DEA-efficient remains in the set of DEA-efficient securities. Figure 2 shows the robustness of the set of DEA-efficient securities through an adjustment factor of 40%.





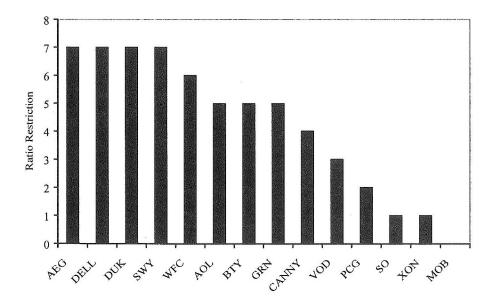
In addition to examination of robustness to unfavorable change in attribute values, it is also important to examine robustness to unfavorable change in weighting restrictions. The weighting constraints presented in equations (6) and (7) state that weights are permitted to be no more than 5 times other weights (.2 <Ratio <5). If these weighting relationships are "tightened," certain securities initially found to be DEA-efficient will leave the set of DEAefficient securities. Table 6 shows several weighting restrictions imposed on the DEA model used for this analysis:

		Table 6	
Listing	of	Weighting	Restrictions

Restriction	Weight Ratio Restrictions
1	(1/4.5) < Ratio < 4.5
2	(1/4) < Ratio < 4
3	(1/3.5) < Ratio < 3.5
4	(1/3) < Ratio < 3
5	(1/2.5) < Ratio < 2.5
6	(1/2) < Ratio < 2
7	(1/1.5) < Ratio < 1.5

Each successive restriction is more binding than the previous one — the upper and lower bounds on each restriction become "tighter," and subsequently, more difficult to remain DEA-efficient. For example, restriction (2) is more restrictive than restriction (1), and so on. Restriction (7) is the most restrictive. For each successive restriction, the analysis is performed to determine which of the securities initially found to be DEA-efficient remain DEA-efficient. The results are presented in Figure 3 below:

Figure 3. Sensitivity Analysis to Restrictions on Weights



For Figure 3, Mobil (MOB) shows no robustness to restriction of weighting ratios. It is important to note that if weighting restrictions were "loosened," the original set of DEA-efficient stocks would remain on the efficient frontier. The reason for this is straightforward: once they are efficient, they will remain efficient if their weights are permitted to take on a wider range of values (McMullen and Strong, 1998).

DISCUSSION OF RESULTS

Tables 3 and 4 show which of the 185 largest market-cap securities are DEA-efficient and near-efficient within the context of this analysis. All of the DEA-efficient stocks have the common characteristic of having their composite inputs being offset by their composite outputs — thereby making them "efficient." Examination of the efficiency frontier in Figure 1 provides further detail into this thought. Duke Energy (DUK) and Canon, Inc. (CAN) can be thought of as securities being on the "southwest" corner of the efficiency frontier. Here, both virtual inputs and virtual outputs are relatively small. In essence, both the cost of obtaining these securities and the associated benefits are relatively low. On the "northeast" end of the efficiency frontier, both virtual inputs and outputs are relatively high. Here, both costs of obtaining the resident securities and their associated benefits are high. Dell Computer (DELL) and America Online (AOL) exemplify this issue. Regardless of where a security resides on the efficiency frontier, however, the common link for all of these is that their virtual inputs are offset by their virtual outputs.

In terms of robustness, a few things are worth noting (the reader is referred to Table 6 and Figures 2 and 3). Dell Computer, AEGON and Safeway all exhibit robustness — they remain DEA-efficient through unfavorable attribute change beyond 40% and with weighting ratios restrictive to 1.5. Other securities, however, are not quite so robust to unfavorable change. Exxon (XON) and the Southern Company (S) are robust to only 5% attribute change and not robust to any tightening of weight restrictions. Mobil Oil (MOB) is robust to a 5% unfavorable change in attribute value, while not being at all insensitive to any type of tightening of weight restrictions. Pacific Gas & Electric (PCG) is robust to only 5% unfavorable attribute change, while being robust to tightening of weighting ratios of only 4.

SUMMARY AND CONCLUSIONS

A technique has been presented which employs Data Envelopment Analysis to select the most desirable securities from a list of the largest market cap stocks. Within the context of this analysis and assumptions, it is shown that of these 185 stocks evaluated, fourteen are found to be relatively efficient, or dominant, while there are four others found to be "near-efficient." One of the advantages of using DEA is that for the DEA-inefficient securities, information is provided disclosing how much reduction of inputs or how much augmentation is needed for these inefficient securities to become DEA-efficient. In other words, DEA can inform the decision-maker which alternatives are consistently the best when several attributes are considered, but it also provides information as to how much improvement is needed for each alternative to become efficient with respect to inputs (input reduction) and outputs (output augmentation).

It must be noted that if a prospective investor were to use this approach for selection of stocks, consideration must be made to weighting issues. The weighting for the initial model (.2 < Ratio < 5) was intentionally generalized. The sensitivity analysis where the weighting schemes were "tightened" was also generalized. Different investors, of course, will have very specific priorities, and weights can be engineered to reflect these priorities. For example, an investor may consider the five-year return more important than the one-year return — the weights can be adjusted to reflect this priority.

The impetus for this research is not necessarily to assist the investor in choosing the best stocks. The motivation for this analysis is to show how Data Envelopment Analysis can be used to assist with the multi-criteria problem of selecting which stocks to invest in. The technique can provide a single, composite score for each alternative, which has simplifying value.

REFERENCES

- Ali, I., (1995). IDEAS 5. 1. Amherst, MA: 1 Consulting.
- Banker, R.D., Charnes, A. & Cooper, W.W. (1984). "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, 30(9), 1078-1092.
- Charnes, A., Cooper, W.W. & Rhodes, E. (1978). "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 2(4), 429-444.
- Charnes, A., Cooper, W., Lewin, A. & Seiford, L. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston, MA: Kluwer Academic Publishers.
- McMullen, P. & Strong, R. (1998). "Selection of Mutual Funds Using Data Envelopment Analysis," *Journal of Business and Economic Studies*, 4(1), 1-12.
- Strong, R.A. (1998). *Practical Investment Management*. Cincinnati, OH: South-Western College Publishing.
- Thompson, R.G., Langemeier, L.N., Lee, C., Lee, E. & Thrall, R.M. (1990). "The Role of Multiplier Bounds in Efficiency Analysis with Application to Kansas Farming," *Journal of Econometrics*, 46(1), 93-108.
- Tobel, E.Z., Value Line Investment Survey, http://www.valueline.com, 1998, New York, NY.
- Wong, Y.B. & Beasley, J.E. "Restricting Weight Flexibility in Data Envelopment Analysis," *Journal of the Operational Research Society*, 1990, Vol. 41, No. 9, pp. 829-835.