Hindawi Publishing Corporation
Advances in Decision Sciences
Volume 2012, Article ID 869128, 9 pages
doi:10.1155/2012/869128

Research Article

A Two-Phase Data Envelopment Analysis Model for Portfolio Selection

David Lengacher and Craig Cammarata

Concurrent Technologies Corporation, 100 CTC Drive, Johnstown, PA 15904, USA

Correspondence should be addressed to David Lengacher, dclengacher@gmail.com

Received 6 October 2012; Revised 30 November 2012; Accepted 18 December 2012

Academic Editor: Henry Schellhorn

When organizations do not have well defined goals and constraints, traditional mixed integer programming (MIP) models are ineffective for portfolio selection. In such cases, some organizations revert to building project portfolios based on data envelopment analysis (DEA) relative efficiency scores. However, implementing the k most efficient projects until resources are expended will not always result in the most efficient portfolio. This is because relative efficiency scores are not additive. Instead, the efficiency of each candidate portfolio must be evaluated against all possible portfolios, making for a computationally intensive task. This paper has two main contributions to the literature. First, we introduce a new DEA-MIP model which can identify the most efficient portfolio capable of meeting organizational goals at incremental resource levels. Second, by utilizing a second-stage DEA model to calculate the relative effectiveness of each most efficient portfolio, we provide managers, a tool for justifying budget increases or defending existing budget levels.

1. Introduction

A critical aspect of management is the decision whereby the best set of projects, or investments, is selected from many competing proposals. In many cases the stakes are high because selecting projects is a significant resource allocation decision that can materially affect the operational competitive advantage of a business [1]. What makes project selection challenging is that the valuation process is oftentimes plagued with high degrees of uncertainty due to long payback periods and changing business conditions. As a result, many researchers have used data envelopment analysis (DEA) as a method by which to evaluate large sets of competing projects [2–6].

DEA was initially developed by Charnes, Cooper, and Rhodes [7] as an efficiency analysis tool and quickly became a popular area in operations research. DEA measures the relative efficiency of decision making units (DMUs) which can represent projects, processes, policies,

or organizations. Although all DMUs must be defined in terms of a common set of inputs and outputs, they do not need to have the same units of measurement. DEA scores efficiency on a scale from 0 to 1 and is thus capable of discriminating among the inefficient units, allowing one to rank projects from most to least efficient.

Although project efficiency scores provide an appropriate basis on which to compare individual projects, the scores cannot necessary be used to assemble most efficient portfolios. The reason behind this is clear. When two or more projects enter a portfolio, we must evaluate the collective inputs and outputs of that portfolio against the collective inputs and outputs of every other possible portfolio, or power set. Only then can we determine whether or not a portfolio is most-efficient. Some researchers have developed DEA models that obviate this computationally intensive task at the cost of introducing a series of subjective decisions to categorize and weight projects. See, for example, Linton et al. [8] and Eilat et al. [9]. However, these approaches require a series of value judgments and preference criteria to facilitate the creation of optimal portfolios. We seek a more objective solution and one with a relatively low burden placed on stakeholders and decision makers.

In our paper, we extend the work of Cook and Green [10] to develop an enhanced DEA mixed integer programming (MIP) model to identify the most-efficient, goal-achieving project portfolios across a range of resource levels. Once these portfolios are determined, we then sum the collective inputs and outputs of each portfolio and treat them as new set of DMUs. Next, we evaluate the relative effectiveness of the set of most-efficient portfolios using an effectiveness-focused DEA model. This postprocessing provides a valuable set of data which can be used to create a trade-off curve of efficiency and effectiveness. As we discuss in detail, this integrated portfolio efficiency-effectiveness curve enables several types of analyses to be conducted depending on the shape of the curve and whether the motive of stakeholders is defensive or offensive in nature. For example, from a defensive perspective, one may use this curve to defend an existing budget level or quantify the degree to which budget cuts or reduced resource levels will affect portfolio efficiency. Alternatively, from an offensive perspective, one may show how specific increases to budget or resource levels will increase portfolio efficiency.

Because these curves are unique to every case and are sometimes nonmonotonic, both of these perspectives of analysis may be possible from the same curve. Our analysis reveals that from certain starting points, a small increase in resource levels (i.e., budget, labor, etc.) can enable the construction of portfolios with increasing efficiency before quickly decreasing at even higher resource levels. Thus, regardless of the political stance of stakeholders, this portfolio efficiency-effectiveness curve can be used to identify win-win bands along the investment continuum. Using our approach, decision makers will be able to identify optimal resource levels, that is, resource levels that achieve maximally efficient portfolios within acceptable ranges of effectiveness.

The rest of this paper is organized as follows. Section 2 discusses the concepts behind and the formulation comprising our DEA-MIP model and post-processing effectiveness model. We then apply our approach to a previously studied data set and examine the results in Section 3. Finally, we discuss the limitations and implications of our approach and future research ideas in Section 4.

2. Material and Methods

In the interest of clarity, we adopt the notation used by Cook and Green [10] to describe our approach. Let us assume a set $P = \{1 \cdots k \cdots |P|\}$ of independent projects where projects

are comprised of a set $O = \{1 \cdots j \cdots |O|\}$ of outputs and a set $I = \{1 \cdots i \cdots |I|\}$ of inputs. Thus, each project k in P is characterized by outputs y_{kj}, which are only possible through the consumption of inputs x_{ki}. Next, let us assume finite amounts of organizational resources, L_i, that can be used to meet the various input requirements of selected projects. Our goal then is to identify a subset $s^* \subset P$ in which we can invest L_i most efficiently. Thus, s^* is the most efficient portfolio.

As mentioned earlier, we can use the following constant returns to scale DEA model to calculate relative efficiency scores for each project in P:

$$\begin{aligned} \text{Maximize} \quad & h_k = \sum_{j \in O} u_j y_{kj}, \quad k \in P, \\ \text{subject to} \quad & \sum_{i \in I} v_i x_{ki} = 1, \\ & \sum_{j \in O} u_j y_{pj} - \sum_{i \in I} v_i x_{pi} \leq 0, \quad p \in P, \\ & u_j, v_i \geq 0, \quad j \in O,\ i \in I. \end{aligned} \tag{2.1}$$

The result is that we will now have basis, h_k, on which to score and prioritize projects. However, as Cook and Green [10] demonstrate, adding efficiencies of those projects in a portfolio will provide an inaccurate measure of the portfolio's true efficiency. Instead, the portfolios collective inputs and outputs must be compared against the power set. It is important to note that the power set includes sets of single projects. Thus, large portfolios must compete against individual projects; a situation which highlights an obvious disadvantage for increasingly larger portfolios. For example, assume we fill a knapsack with the most efficient projects first and continue filling it in descending order of efficiency. Naturally then, one can see how larger and larger portfolios necessarily lead to a general decrease in efficiency, while at the same time total output (effectiveness) increases.

Things become more complicated when fixed resources, L_i, enter the picture. For example, assume we have a budget that allows us to choose between portfolio A, comprised of the single most and least efficient projects, or portfolio B, comprised of the second most efficient and second least efficient project. Which one should we chose? The answer is not obvious because it is inappropriate to compare portfolios based on the efficiency scores of their members. Thus, we need a model that combines selection and efficiency evaluation, a knapsack-DEA model. Using the same notation, the following linear program by Cook and Green [10], whose work stemmed from Oral et al. [11], solves this problem:

$$\begin{aligned} \text{Maximize} \quad & h_k = \sum_{k \in P,\ j \in O} a_{kj} y_{kj}, \\ \text{subject to} \quad & \sum_{k \in P,\ i \in I} b_{ki} x_{ki} = 1, \\ & \sum_{j \in O} u_j y_{pj} - \sum_{i \in I} v_i x_{pi} \leq 0, \quad p \in P, \end{aligned}$$

$$\begin{aligned}
&\sum_{k\in P} c_k x_{ki} - l_i = L_i, \quad i \in I,\\
&(1-c_k)x_{ki} + Mc_k + Md_k \geq l_i + \frac{1}{M}, \quad k \in P,\ i \in I,\\
&\sum_{i\in I} d_{ki} \leq |I| - 1, \quad k \in P,\\
&c_k, d_{ki} \in \{0,1\}, \quad k \in P,\ i \in I,\\
&a_{kj} \geq 0, \quad k \in P,\ j \in O,\\
&a_{kj} \leq Mc_k, \quad k \in P,\ j \in O,\\
&u_j \geq a_{kj}, \quad k \in P,\ j \in O,\\
&b_{ki} \geq 0, \quad k \in P,\ i \in I,\\
&b_{ki} \leq Mc_k, \quad k \in P,\ i \in I,\\
&v_i \geq b_{ki}, \quad k \in P,\ i \in I,\\
&v_i \leq b_{ki} + M(1-c_k), \quad k \in P,\ i \in I,\\
&M > 0.
\end{aligned} \tag{2.2}$$

Here, c_k is 1 if project k is included in subset s, and 0 otherwise. The slack variable l_i represents the remaining or unused portion of resource L_i. To make the problem linear, two variable substitutions are made: $a_{kj} = c_k u_j$ and $b_{ki} = c_k v_i$. Finally, the Big M formulation is used to enforce the following constraint, which ensures the portfolio under evaluation cannot be made larger because of insufficient resources to add any remaining projects:

$$(1-c_k)x_{ki} + Mc_k + Md_k \geq l_i + \frac{1}{M}. \tag{2.3}$$

As Cook and Green [10] highlight, the key innovation in this model is the discovery of a redundancy in the constraints of prior models which included comparing each candidate portfolio against the power set. As it turns out, because the efficient frontier of the power set will always be defined by singleton sets, the best nonsingleton portfolios can do is to lie on the frontier defined by singleton sets. Thus, because the constant returns to scale DEA model measures radial distance between a DMU and this frontier, the result is that candidate portfolios need only to be compared against singleton sets, and not the power set. The outcome is an algorithm that is much less computationally intensive.

It is important to note that (2.2) will identify not simply the most-efficient portfolio from the power set, but instead the most-efficient portfolio that fully utilizes all of resources L_i, to the point where no other projects can be added due to insufficient remaining resources.

By adding the following constraints to (2.2), we can increase the flexibility of the model in several important ways:

$$\sum_{k\in P} c_k y_{kj} \geq G_j, \quad j \in O,$$
$$\sum_{k\in P} c_k y_{kj} = CO_j, \quad j \in O. \tag{2.4}$$

First, recall that each project k in P is characterized by outputs y_{kj}, which are only achievable through the consumption of inputs x_{ki}. Next, let us assume that G_j represents organizational goals to meet or exceed cumulative output j. Then, the first constraint ensures that only portfolios whose cumulative outputs meet or exceed stated goals are feasible and can be evaluated.

Before discussing the value of the second constraint it is important to consider the question of how best to graphically evaluate the efficiency and effectiveness of portfolios. Solving (2.2) for different levels of L_i will produce a set of most-efficient portfolios and their corresponding efficiency scores, which may increase or decrease in L_i. To calculate effectiveness scores of each most-efficient portfolio, we must first recognize that while efficiency is a measure of total output/total input, effectiveness is a measure of output only. Next, because portfolio output is multidimensional, we utilize a DEA model like (2.1) to provide a frontier-based measure of total output (effectiveness). We follow the methods used by Chang et al. [12] and Tsai and Huang [13] to calculate relative effectiveness scores using DEA. The only difference between their approach and traditional efficiency-focused DEA models is the replacement of all inputs, x_{ki}, with a single vector of 1 s. The result produces relative effectiveness scores. Thus, this second constraint does not affect the solution of (2.2). Instead, it simply allows us to automatically generate cumulative outputs, CO_j, of each solution which the effectiveness-focused DEA model needs.

Using the same data set of 37 projects from the iron and steel industry that Cook and Green [10] and Oral et al. [11] examined, we apply our two-phase approach for generating a portfolio efficiency-effectiveness curve across 16 incremental resource levels. The projects are defined by a single input and five outputs which represent the project resource requirements and expected benefits, respectively. Project resource requirements ranged from 28 to 96 and all 37 projects can be developed for a total of 2515 resource units. Our test range spanned from $L_1 = 200$ to 1970 resource units, in increments of 110 units. In addition, we specified a minimal goal, $G_1 \geq 50$, for our organizational goal constraint. It is important to note that the previous authors only investigated the solution for a single resource level of $L_1 = 1000$. This produced a portfolio of 16 projects. Upon inspection, these 16 projects happened to also be the 16 most efficient projects according to the individual DEA efficiency scores. Because this coincidence was not discussed in Cook and Green [10], the reader of that study may incorrectly conclude that this is always the case and believe the whole idea of finding the most efficient portfolio is rather trivial, simply continuing to add the next most efficient project to the portfolio. One of the contributions of our work is to illustrate situations where this coincidence does not occur.

3. Results

Our analysis revealed there is a strong inverse relationship between effectiveness and efficiency in the data set (see Figure 1). Starting from a resource level $L_1 = 200$, the model

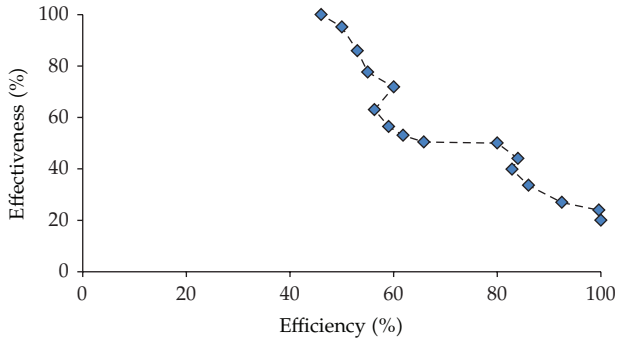

Figure 1: Portfolio efficiency-effectiveness curve.

chose a portfolio of four projects with a relative efficiency of 98% and a relative effectiveness of 18%. The last resource level tested, $L_1 = 1970$, generated a portfolio of 29 projects with a relative efficiency of 44% and relative effectiveness of 100%. Throughout the 16 test points we witnessed some projects entering and leaving s^* several times. This behavior clearly demonstrates how determining the most efficient portfolio that utilizes all resources maximally cannot be achieved using a strictly additive process like continuously adding projects to a portfolio based on their individual efficiency scores.

Of the 16 portfolios generated, portfolio effectiveness increased with every increment in L_1. This behavior is expected, as effectiveness DEA models focus solely on outputs and, thus, the most-effective portfolio will always be the one that includes all projects [12]. The most noteworthy aspect of this curve is that of 16 portfolios, two are efficiency dominated by the portfolio immediately succeeding (above) it. These investment levels can be viewed from different perspectives. First, we can reasonably categorize these points as generally undesirable when compared to the solutions immediately succeeding them. Thus, if given even some small amount of control of our own resource levels, we should try to avoid these points. Alternatively, if we find ourselves currently at these points, we may view our situation as a great opportunity; a situation in which a win-win scenario (an increase in both efficiency and effectiveness) is within immediate reach.

It is important to note that the efficiency oscillations in this type of curve will increase significantly when relatively efficient projects also happen to be the more resource-intensive projects (i.e., more expensive). To illustrate this point, we analyzed six notional water conservation projects where the most efficient project is also the most expensive one. It is easy to see that the switch-back in the portfolio efficiency-effectiveness curve is more pronounced than those seen in Figure 1. Here we find that as soon as the available resource level reaches a point allowing us to afford the most expensive and most efficient project, we create a portfolio containing only that project. The result is that this new portfolio will be more efficient than all prior portfolios generated at lower resource levels (see Figure 2).

Next, as we continue to solve the two-stage model and increase L_i, the curve begins to shift up and left once again as we are forced to add less efficient projects back into our portfolio in order to maximally expend our resources (See Figure 3). As before, the most-effective portfolio will always be the one that includes all projects.

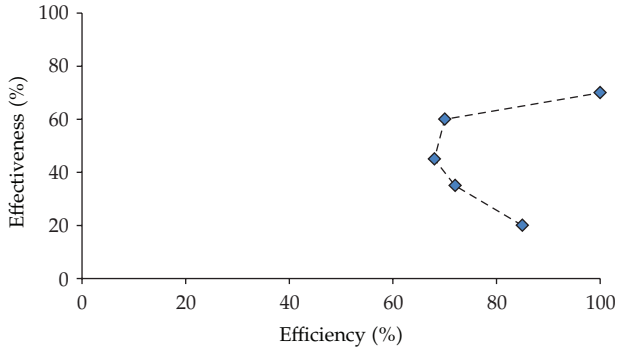

Figure 2: Case in which the most expensive project is also the most efficient.

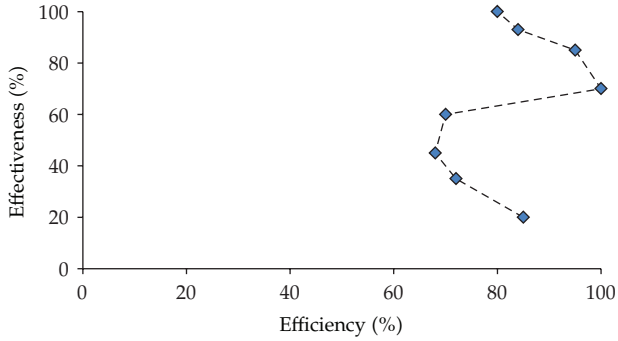

Figure 3: At higher resource levels, less efficient projects are added back to the portfolio.

4. Discussion

The primary limitation of our approach is that it assumes a decision maker has some ability to affect the available resource levels at his disposal for a particular business unit. For this reason, our approach may not be applicable in some situations. For example, if a manager has no ability to increase his budget and faces no risk of budget decrease, then he may find no value in examining a portfolio efficiency-effectiveness curve. Instead, he may simply want to solve (2.2) for a single resource level to determine how best to utilize his fixed resources.

Consequently, our approach may only be applicable at the highest levels of organizational management, environments where decision makers must divide a single pool of resources among multiple business units with independent goals. In this situation, each business unit is tasked with solving its own portfolio optimization problem. This provides a particularly interesting application of our approach, because instead of a single portfolio efficiency-effectiveness curve, decision makers are now faced with multiple curves, one for each business unit.

This set of optimization problems can quickly become complex because fixed resources create decisions that are interdependent; the allocation of resources to one business unit will depend on the combined resources allocated to all other business units. This presents a parallel optimization problem. Thus, we may need another tier of optimization modeling which can consider all the feasible combinations of spending a single budget across independent initiatives, each of which represents a portfolio optimization problem.

Another limitation of our approach is that investment decisions involving project proposals with more than one resource requirement will not result in a continuous portfolio efficiency-effectiveness curve. This outcome occurs because all combinations of resource levels, for each resource type, must be tested. Thus, one must solve for all feasible portfolios that exhaust each resource type independently of the other resource types. This may detract from one of the attractive features of our approach, that the analysis results are simple to comprehend.

5. Conclusion

Avkiran et al. [14] stressed the importance of "pushing the DEA research envelope" by finding new application areas for DEA. By focusing on the ubiquitous organizational task of building project portfolios and quantifying them in terms of efficiency and effectiveness, our aim is to respond to their call to action. This paper expands the resource allocation literature on the use of DEA in the context of tradeoffs between efficiency and effectiveness across a range of resource allocation levels. Our approach is most applicable in hierarchical organizational environments in which a unit (a) can influence, to some extent, its own allocation of resources from a parent unit, or (b) acts as a central unit that controls the resource allocations of a set of units. Our two-phase approach consists of extending an MIP-DEA model and adding a post-processing DEA model to create tool allowing decision makers to quickly identify win-win situations along the investment continuum with respect to efficiency and effectiveness.

References

[1] C. T. Chen and H. L. Cheng, "A comprehensive model for selecting information system project under fuzzy environment," *International Journal of Project Management*, vol. 27, no. 4, pp. 389–399, 2009.

[2] A. Asosheh, S. Nalchigar, and M. Jamporazmey, "Information technology project evaluation: an integrated data envelopment analysis and balanced scorecard approach," *Expert Systems with Applications*, vol. 37, no. 8, pp. 5931–5938, 2010.

[3] J. A. Farris, R. L. Groesbeck, E. M. van Aken, and G. Letens, "Evaluating the relative performance of engineering design projects: a case study using data envelopment analysis," *IEEE Transactions on Engineering Management*, vol. 53, no. 3, pp. 471–482, 2006.

[4] H. K. Hong, S. H. Ha, C. K. Shin, S. C. Park, and S. H. Kim, "Evaluating the efficiency of system integration projects using data envelopment analysis (DEA) and machine learning," *Expert Systems with Applications*, vol. 16, no. 3, pp. 283–296, 1999.

[5] M. A. Mahmood, K. J. Pettingell, and A. I. Shaskevich, "Measuring productivity of software projects: a data envelopment analysis approach," *Decision Sciences*, vol. 27, no. 1, pp. 57–80, 1996.

[6] T. Sowlati, J. C. Paradi, and C. Suld, "Information systems project prioritization using data envelopment analysis," *Mathematical and Computer Modelling*, vol. 41, no. 11-12, pp. 1279–1298, 2005.

[7] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[8] J. D. Linton, S. T. Walsh, and J. Morabito, "Analysis, ranking and selection of R&D projects in a portfolio," *R&D Management*, vol. 32, no. 2, pp. 139–148, 2002.

[9] H. Eilat, B. Golany, and A. Shtub, "Constructing and evaluating balanced portfolios of R&D projects with interactions: a DEA based methodology," *European Journal of Operational Research*, vol. 172, no. 3, pp. 1018–1039, 2006.

[10] W. D. Cook and R. H. Green, "Project prioritization: a resource-constrained data envelopment analysis approach," *Socio-Economic Planning Sciences*, vol. 34, no. 2, pp. 85–99, 2000.

[11] M. Oral, O. Kettani, and P. Lang, "A methodology for collective evaluation and selection of industrial R&D projects," *Management Science*, vol. 37, no. 7, pp. 871–885, 1991.

[12] P. L. Chang, S. N. Hwang, and W. Y. Cheng, "Using data envelopment analysis to measure the achievement and change of regional development in Taiwan," *Journal of Environmental Management*, vol. 43, no. 1, pp. 49–66, 1995.
[13] K. C. Tsai and Y. T. Huang, "Using data envelopment analysis (DEA) method to establish the grading framework in college english composition class," *Education Technology Letters*, vol. 1, no. 1, pp. 15–21, 2011.
[14] N. K. Avkiran and B. R. Parker, "Pushing the DEA research envelope," *Socio-Economic Planning Sciences*, vol. 44, no. 1, pp. 1–7, 2010.