Abstract

The concepts of Data Envelopment Analysis (DEA) are reviewed briefly in connection with the problem of assessing the impact of resources. The problem is complex, as there are many measures of impact, and the situation of every border station or port of entry is in some ways unique. DEA addresses that uniqueness by applying techniques of Linear Programming to adjust the relative weights of both impact measures and resources used, to “make each station as efficient as possible.” The resulting ratios of weights for impacts, to weights of costs, provide an indication of how well that station might use the resources to achieve its goals. It is proposed that the data set generated in this analysis might be used effectively to drill down and assess the impact of each specific resource, in a way that reflects the nuance and complexity of border security.

1 Introduction

When multiple kind of resources, in quantities $x_1, \cdots, x_R$ can be brought to bear on border security, it seems natural to ask which resource, $r^*$, “has the most impact” on the security of the border. In this note we review some conventional ways to address this question, and
propose a way in which Data Envelopment Analysis (DEA) might provide some new insights into the question. We begin by reviewing the basic structure of the problem, and considering a canonical econometric approach to the problem. We then, in Section 2 discuss an approach using DEA. We conclude with some discussion of ways in which this work might be carried forward (Section 3).

There are four important complications to the apparently simple question of impact.

First, the impact of any specific resource may depend on how it is used. Let us suppose that we want to know the impact assuming that each resource is used as it is presently, so that current and historical data can provide a guide.

Second, there are several ways to measure the security of the border (for example, by reductions in the flow of persons (by risk categorization!), of narcotics (by type!), of counterfeit goods (by economic impact!), etc. Let us generically designate those measures as $y_1, \cdots, y_I$, where $i = 1, \cdots, I$ labels the distinct kinds of impact.

Third, flows across the border occur at places (Ports of Entry (POE)) or in areas (sectors, zones, etc. between POEs). A resource may be applied specifically to one place, which may be denoted as $x_r^{(p)}$, the quantity of resource $r$ applied to place $p$. Or a resource may act globally (for example, a change in administrative policies). The quantity of a resource that acts globally will be denoted here by $x_r^{(0)}$.

Fourth, the impact of an increment in any specific resource, $\delta x_r^p$ may depend not only on how it is used (which we have taken off the table) and on “environmental factors” (e.g., the terrain, the weather, the global economy), but also on the level of other resources. A simple example illustrates the problem. If drones are deployed, they require operators and spotters, who control them, and scan the transmitted images. Adding more operators, without drones for them to fly, may provide no improvement at all. Mathematically, there is a non-linear relation which might be expressed as shown in Equation 1. Here $n_{o,d}$ represent the numbers of operators and drones, respectively. The parameter $a$ represents the value of the right mix, while the parameter $b$ represents the cost of having an imbalance in either direction.
\[ y = an_d - b|n_d - n_o| \]  

1.1 Multivariate Production Frontier

The canonical approach to investigating the relation between resources and outcomes is to assign an overall value function \( V(y) \), where \( y \) is the vector of outcomes. Assuming the existence of a production function: \( y = P(x) \), where \( x \) is the vector of resources, the overall value is: \( V = V(P(x)) \). The incremental value of any specific resource is (assuming for the moment that both \( V \) and \( P \) are differentiable functions of their arguments):

\[ v_j(x) = \sum_i (\partial v_i/\partial y_i)(\partial y_i/\partial x_j) \]

Determination of the function \( P \) requires substantial amounts of data, and the selection among several possible model functions. Linear functions, although popular in theoretical analyses, are most unlikely to accurately reflect the realities at the border. This is even true of the relation between the value function \( V \) and the several indicators of impact, \( y \), as national policies and common sense suggest that larger problems loom more importantly in the assessment of overall border security.

That having been said, the purpose of this brief note is rather to explore the perspective of DEA, and to ask whether it might provide another way of thinking about the issues.

2 Analysis

Data Envelopment Analysis is a powerful mathematical technique that can be used to assess the “efficiency” of many different organizations (called “Decision Making Units” or DMUs) that work towards similar goals, but do so in differing operational environments. Examples would be the hospitals in a large system, the public schools in a large system, or a large collection of libraries. The theory was laid out in Charnes et al. (1978) at a time when
the computations involved were quite challenging. In the intervening years, the range of applications has grown enormously, as reviewed in Cook and Seiford (2009); Reichmann (2004); Cooper et al. (2011, 2007). In addition, purpose-built software has become available, costing less that $1,000 to own (for academic use).

Briefly, the concept of DEA is that each DMU may assign its own relative weights to the several measures of impact, \( y_i \) and its own relative costs to the several resources \( x_j \). Let us say that the weights assigned at position \( p \) are given by \( u_i^{(p)} \), \( c_j^{(p)} \), where \( u \) represents the local utility function, and \( c \) represents the local cost. DEA then permits each DMU, \( p \), to adjust the sets of weights so that the apparent cost-effectiveness of that DMU is as large as it can be. The apparent cost effectiveness is measured as given in Equation 3.

\[
E_{\text{apparent}}^{(p)} = \frac{\sum_i u_i^{(p)} y_i^{(p)}}{\sum_j c_j^{(p)} x_j^{(p)}}. \tag{3}
\]

It is apparent that the approach can be “gamed” by finding the resource that is least used, and the impact that is highest, and setting all other weights to zero. This is prevented by imposing constraints on the ratios of the costs, \( c_j : c_j' \) and of the partial utilities, \( u_i : u_i' \). Absent any other information, these become complicated constraints ensuring that no DMU is using weights too far from those used by the others. When there is other information available on costs and on benefits, these weights may be constrained to be, e.g., within a factor of two, or or four, of their overall agreed-upon values.\(^1\) This approach has been used by the author and a colleague, in studies of academic libraries, Shim (2000, 2003); Shim and Kantor (1999, 1998).

The results are then used in a sophisticated exploration of whether each DMU is “doing the best that it can,” in getting impact from its resources. To assess this, DEA plots the weighted cumulated impact, against the weighted cumulated cost for all of the DMUs, using

\(^1\)The resulting optimization problem can be solved by techniques of Linear Programming, and there are several packages available, some of which specifically focus on Data Envelopment Analysis. Examples are found at: https://deaos.com; http://www.deafrontier.net/deafree.html; and http://opensourcedea.org/. Of these, the author has examined only deafrontier.
the same weights that make a specific one, say $K$ “look best” (that is, maximize the apparent effectiveness $E^{(K)}_{\text{apparent}}$). For each DMU there is a separate plot. And for each DMU, one of two things may happen: (a) the DMU is on the Pareto frontier\(^2\) of the plot (b) it is not on the frontier, but falls below that frontier. In case (a) we say that the DMU is “100% efficient.” In case (b) we select one of three ways to assess efficiency. Each results in choosing a different point along the frontier at which to evaluate the “best possible efficiency,” $E^{(p)}_{\text{ideal}}$. The ratio of Equation 4 is taken as the “DEA measure of relative efficiency, for the DMU.”

$$\frac{E^{(p)}_{\text{apparent}}}{E^{(p)}_{\text{ideal}}}$$ (4)

Assuming that all the technical hurdles are cleared: the values of $x, y$ have been gathered for every position, $p$, and the values of the weights that make each DMU appear at its best, we can return to the question of assessing the “impact of a resource.” We do not propose examining the overall efficiency of each DMU, but concentrate rather on the range of impact-cost ratios for each resource, obtained by considering all the DMUs and all the different measures of impact. Let us denote those values by $r^{(p)}_{i,j} = \frac{u^{(p)}_{i}}{c^{(p)}_{j}}$. This is a set of real-valued positive numbers, indexed by three categorical labels: $p$, labeling the DMU; $i$, labeling the specific measure of impact; and $j$, labeling the specific resource.

This array of data, if it ever becomes available, can be used in many ways. A centralized decision maker might rescale costs, to reflect the total cost of providing and maintaining the resource (for example, it may prove difficult to recruit new personnel to remove border stations, above and beyond the salary expense.) Similarly, the several impacts might be weighted to reflect changes in national policy, and the weight assigned to a specific DMU might reflect any complex mix of considerations.

\(^2\)This is the upper left boundary of the convex hull of the set of points in the cost-impact plane for all the DMUs, including the origin.
3 Conclusions and Discussion

In sum, we have presented a way in which the concept of Data Envelopment Analysis can provide a much richer and more nuanced picture of the impact of any specific resource. This approach uses a sophisticated, but computationally tractable method to produce a three-dimensional data cube indicating the DEA-based: impact of a resource on a measure of impact, specified as to location. This research, to our knowledge, cannot yet be implemented because the data are not publicly available. It is possible that US Border Protection could assemble all of that data, and conduct an internal review of the considerations presented here. We note that the data might also support a conventional econometric analysis, but propose that the DEA approach better accommodates the manifest differences among the diverse zones and ports that must be protected.

With any economic analysis there is the danger that the hidden computations may produce findings that are politically unpalatable, not because they are wrong, or treat one or another stakeholder unfairly, but simply because they are presented as a “black box.” In an era of increasing skepticism, transparency must be a consideration in analysis and decision making. To that end, research on the visual presentation of this kind of data could be very helpful. To recall, the data are “ratio-scale data,” labeled by three categorical variables, position, impact-measure, and resource. Because we live in a three-dimensional world, ingenuity is required to present relations of this complexity.

The path forward. We propose that the possibility of gathering the requisite data, and the assessing the suitability of a DEA-based analysis be considered as a research direction, by the Department of Homeland Security, for CBP, or for other aspects of its mission.

Acknowledgments. This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2015-ST-061-BSH Subcontract: R-17-0050 The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. I thank Dennis Egan for ongoing
leadership of this research project.

References


