

4-1-2003

Extending Productivity Research Frontiers: DEA Resource of Datasets and Errata

Timothy R. Anderson
Portland State University, tim.anderson@pdx.edu

Paul Rouse
University of Auckland

Rafael Borja
Portland State University

Ivan Patricio Hernandez
Portland State University

Fabricio Tobar
Portland State University

See next page for additional authors

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/etm_fac

 Part of the [Engineering Commons](#)

Citation Details

Anderson, Timothy R.; Rouse, Paul; Borja, Rafael; Hernandez, Ivan Patricio; Tobar, Fabricio; and Setiowijoso, Liono, "Extending productivity research frontiers: DEA resource of datasets and errata" (2003). Engineering and Technology Management Faculty Publications and Presentations. Paper 20. <http://archives.pdx.edu/ds/psu/9261>

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Engineering and Technology Management Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Authors

Timothy R. Anderson, Paul Rouse, Rafael Borja, Ivan Patricio Hernandez, Fabricio Tobar, and Liono Setiowijoso

Extending Productivity Research Frontiers:

DEA Resource of Datasets and Errata

Timothy R. Anderson¹, Paul Rouse², Rafael Borja¹, Ivan Patricio Hernandez¹, Fabricio Tobar¹, and Liono Setiowijoso¹

¹Dept. of Engineering & Technology
Management
Portland State University
Portland, OR 97207-0751
USA
tima@etm.pdx.edu

²Department of Accounting & Finance
The University of Auckland
Auckland
NZ
p.rouse@auckland.ac.nz

Abstract

Website databases provide the opportunity to fundamentally change the way research and knowledge is advanced and disseminated. This paper describes a website, <http://www.etm.pdx.edu/dea/dataset/>, focused on Data Envelopment Analysis and DEA researchers.

1. Introduction

Web-enabled databases provide the opportunity to fundamentally change the way research and knowledge is advanced and disseminated. Typical of these are the Statistical Reference Datasets provided by the National Institute of Standards and Technology (NIST) [1] that provide both real and constructed data for assessing the numerical accuracy of statistical software. Other more unusual datasets include the International Archive of Education Data [2] and the Consortium for Research on Employee Benefits [3]. In this note, we describe a web-site focused on Data Envelopment Analysis (DEA) and DEA researchers.

Since its inception in 1978, a considerable number and variety of DEA models and applications have appeared in the literature [4] with a tendency to provide new datasets for each model and / or application. Consequently, the number of datasets has proliferated. We believe that standard datasets could be used for many papers if they were made available for this purpose.

The ability to replicate is fundamental to the methodological requirement for 'empirically testable' theories e.g. Popper [5]. This process of replication is improved considerably if (i) standard datasets are available as reference benchmarks and (ii) anomalies can be investigated and reported more easily.

The power and ease of use of web-based technology opens up new opportunities for both consolidating and expanding the DEA knowledge base. The ability to consolidate knowledge derives from its archiving role, search tools and linkages to bibliographic details. Expansion of knowledge can be facilitated by access to well-specified datasets tailored to research objectives and model testing. For example, a particular model can be tested using several different datasets, or different models can be tested using the same dataset. Alternatively, new software can be benchmarked against datasets with specified performance characteristics.

Given this need, we have developed a web site (<http://www.etm.pdx.edu/dea/dataset/>) with a mission to provide the necessary flexibility and capability to respond to the needs of the DEA community. The site aims to provide the following:

- a comprehensive range of datasets with guidelines for particular applications, which encompass public and private organizations, specialist applications, small through to large datasets, datasets suitable for classroom demonstrations or exercises.
- a primer that describes each data set that can be a much more comprehensive description than in a typical journal paper;
- sophisticated search enablers i.e. keywords for classifying datasets and enabling selective searches. For example, datasets suitable for non-discretionary models, health applications, panel data for window analyses.
- additions or enhancements to a data set. It is often the case that additional research or new insights enable an 'older' data set to be expanded or otherwise improved.;
- datasets that are especially suited for particular model testing. This is an ongoing project to identify datasets that display 'good' behavioral features for constant or variable returns to scale, strong or weak disposability, categorical or ordinal, changes in technology;
- a history of datasets linked to bibliographies to enable a lineage to be identified. Our goal is not to establish a comprehensive bibliography as such but to provide the references for the data set and a "pedigree" or "genealogy" of the use of those datasets.
- benchmark datasets for software testing with results if published, or made available by researchers;
- *errata* cross-referenced to bibliographies and datasets;

Improved search options allow for finer grained inquiries and tailored responses to user needs. Drop down menus and available databases of names and journals have been incorporated to make the site as friendly as possible, e.g. the entry screen makes use of over 500 names from the DEA research and practitioner community. Wherever possible, data entry has been made easier through the use of such related databases.

In order to succeed, the site must be accepted by researchers and practitioners as worthwhile. Previous researchers have recognized the critical need for a standard data set repository and have taken first steps toward this, but have not been able to sustain the necessary effort. Beasley [6] appears to have been the first to provide datasets from DEA applications, and his site contains four published datasets. Other sites include those of Warwick University [7], PARN [8], and Holger Scheel [9].

As part of our program to achieve a successful implementation, the site is to be developed in close consultation with the wider DEA community by means of conference panel presentations and more detailed briefings with editors of relevant journals and leading members of the research and practitioner groups. To date, endorsements have been obtained from *Journal of Productivity Analysis* and *IEEE Transaction on Engineering Management*. Ongoing efforts are directed at adding other high-level mainstream journals to these. A number of leading DEA researchers have signaled their support and a process is under way to establish formal advisory links between these groups and the site.

2. DEA Dataset on the Web

Currently all datasets available on the web are stored in the format in which they were uploaded. Based on feedback from potential users at the North American Productivity Workshop, we recommend uploading files as Microsoft Excel spreadsheets. Advantages to the spreadsheet format are the wide availability of spreadsheets, the easy manipulation of the

format to fit the requirements of the particular analysis software used, and the additional opportunity for documentation available in the spreadsheet format.

Most DEA software can accommodate spreadsheets or even require the data to be in spreadsheet format. Furthermore, spreadsheets allow easy export of data in comma or tab delimited format for software that requires the data to be in text format.

The database system developed for storing DEA datasets allows researchers to search for appropriate data sets using a wide variety of criteria. For example, algorithm or software developers might want to examine data sets with a large number of dimensions in terms of the number of DMUs, inputs, or outputs. Other researchers may wish to focus on reproducing classic results or viewing the available datasets in a particular application area such as healthcare. The screen for doing searches on The DEA Dataset Repository is illustrated in figure 1.

3. A Call for Datasets

This DEA resource of datasets and errata provides an avenue for the DEA community to advance up a level by systematically structuring its learning both in terms of past endeavors and future research. The advantages are many: multi-search enablers, easy upload and download, archival provision, real and constructed datasets. Future plans include the provision of a lineage for each dataset, dataset behavior features (e.g. VRS, weak disposability), standard versus nonstandard classifications, small to very large numbers of DMUs. We encourage the DEA community to use this service as a platform for advancement as well as a means for archiving research sources.

There are also potential research opportunities inherent in the datasets themselves. What makes a good dataset and how can we tell a good dataset from a bad one? Are there distributional properties of some empirical datasets that lend themselves to better statistical

inferences and / or bootstrapping? Is there any commonality of measures used in particular application areas?

We end with a plea to the DEA community to support our endeavor especially in the form of datasets. We need to build a critical mass and as they say, all contributions are gratefully received!

Please enter the criteria that you want to Search for:

Number of DMU's	20	<= # DMU <=	200
	<input type="text"/>		<input type="text"/>
	<input type="button" value="Search"/>		
Number of Inputs	3	<= # Input <=	5
	<input type="text"/>		<input type="text"/>
	<input type="button" value="Search"/>		
Number of Outputs	3	<= # Output <=	5
	<input type="text"/>		<input type="text"/>
	<input type="button" value="Search"/>		
Analysis Technique	CRS (CCR) <input type="button" value="Search"/>		
	<input type="button" value="Search"/>		
Application Category	Health Care <input type="button" value="Search"/>		
	<input type="button" value="Search"/>		
Author	Rhodes, E. L. <input type="button" value="Search"/>		
	<input type="button" value="Search"/>		
Journal	European Journal of Operational Research <input type="button" value="Search"/>		
	<input type="button" value="Search"/>		
Year	Before 1984 <input type="button" value="Search"/>		
	<input type="button" value="Search"/>		
Between years	Select a Year	<= year <=	Select a Year
	<input type="text"/>		<input type="text"/>
	<input type="button" value="Search"/>		
Dataset & Application description containing the word	<input type="text"/>		
	<input type="button" value="Search"/>		

Figure 1. DEA Dataset Repository on the web

References

1. NIST, *Statistical Reference Datasets*. 1999, NIST, <http://www.itl.nist.gov/div898/strd/>.
2. Statistics, N.C.f.E., *International Archive of Education Data*. 2000, <http://www.icpsr.umich.edu/IAED/>.
3. *Consortium for Research on Employee Benefits*. 2000, <http://www.ifebp.org/recreb.html>.
4. Seiford, L.M., *A bibliography for Data Envelopment Analysis (1978-1996)*. *Annals of Operations Research*, 1997. **73**: p. 393-438.
5. Popper, K.R., *Conjectures and Refutations*. 1963, London: Routledge and Kegan Paul.
6. Beasley, J.E., *OR-Library*. 1994, <http://www.ms.ic.ac.uk/info.html>.

7. Emrouznejad, A., *Ali Emrouznejad's DEA Home Page*. 2000, <http://www.warwick.ac.uk/~bsrlu/findex.htm>.
8. University, O., *PARN - Productivity Analysis Research Network"Other Information"*. 2000, School of Business and Economics, Department of Management, <http://www.sam.sdu.dk/parn/othinfo.htm>.
9. Scheel, H., *Holger Scheel's Data Envelopment Analysis Page*. 2000, <http://www.wiso.uni-dortmund.de/lsg/or/scheel/doordea.htm>.