ASSESSING LIBRARY PERFORMANCE IN THE SHORT AND LONG RUNS: EFFICIENCY ANALYSIS AND EMPIRICAL APPLICATION

Frederico A. de Carvalho
UFRJ (the Federal University at Rio de Janeiro)

Marcelino José Jorge
LAPECOS - FIOCRUZ (Oswaldo Cruz Foundation)

Marina Filgueiras Jorge
INPI (the National Institute of Industrial Property)

1 Introduction

This paper presents an optimization approach to assessing library performance from an Efficiency Analysis standpoint. Concerning the empirical illustration, the paper employs data from a sample pertaining to a public system of academic libraries. Our approach combines in a simple way efficiency scores computed from the estimation of selected Data Envelopment Analysis (DEA) models and a long run evaluation provided by Markovian analysis. Managerial implications are mentioned throughout the text.

2 Background

The proposed approach relies essentially on the application of the so-called Efficiency Principle to a public university system of libraries. Following the literature, "organization" may be taken quite broadly as meaning both public and private entities, and even nonprofit ones among them (see Carvalho *et al*., 2012).

Data Envelopment Analysis (DEA)

The Efficiency Principle simply states that, whenever a production unit uses the same resources but yields greater quantities of output than another unit, it should be considered "relatively more efficient" (i.e., relative to one another). Analogously it should be considered "relatively more efficient" if it uses fewer resources and yields the same output. From a modeling standpoint these properties correspond to evaluating a library unit in terms of its position *vis à vis* an adequately defined and computed "efficiency frontier", that is, the *locus* of all "equally best productive combinations of inputs and outputs". Once identified the frontier, the performance of a specific library system may be evaluated by assessing the relative position of its component units relatively to each other and to the frontier.

There is an established body of knowledge – namely, Data Envelopment Analysis (DEA), a class of mathematical programming models – with a now long tradition (see references in Carvalho *et al*., 2012) of being applied to a broad range of situations involving the analysis of production frontiers in a multi-unit, multi-input and multi-output framework where parametric restrictions are absent. The so called nonparametric models of frontier adjustment, such as DEA, represent the efficiency frontier as the best observed practices, that is, as the maximum output obtained from an input bundle when considering all the empirically observed organizational units in the population studied.
In applied work DEA has been used to evaluate several types of organizations, such as libraries (see references in Carvalho *et al*., 2012), industrial plants, bank branches,

educational systems and/or units, and hospitals, all properly understood as examples of "complex organizations".

Efficiency Analysis in the long run

In a seminal methodological paper Tulkens and Vanden Eeckaut (1995) describe and explain the main issues relating to the role of time in nonparametric efficiency analysis, especially in what concerns alternative ways to accommodate empirical information into reference production sets that will be submitted to efficiency computations. Their presentation depicts different approaches to panel data efficiency analysis. Of particular interest here (see Table 1) is their classification (p. 478-480) of the variety whereby time dimension may be treated.

There is some previous literature focussing the long run in (in)efficiency analysis (see references in Carvalho *et al*., 2012). Some authors employed econometric techniques specifying a lag structure so that (long run) equilibrium may be discussed with appropriate difference equations. More close to the approach adopted here, the paper by Wang and Huang (2007) introduced a two-state Markov Chain model leading to the estimation, for each DMU, of its efficiency status as specified in their equation (2.12) (p. 1307).

Using results from finite ergodic Markov chains (Kemeny and Snell, 1972, p. 130-131), and assuming one (estimated) aggregate transition matrix is available, this paper computes the fundamental matrix and the long run distribution of the "system" (the set of DMUs) between the two states.

3 Method

Our proposed procedure consists of three steps. The first two steps – involving the computation of efficiency scores and of operational plans in turn – are typical in many applications of Data Envelopment Analysis to performance assessment. The third step, a novel one, incorporates the "structural" long run assessment of efficiency.

Data collection

The case is summarized in Table 1 following the Tulkens and Vanden Eeckaut (1995) framework.

Table 1 – Summary on case study

Case (DMUs)	Number of DMUs	Number of variables	Time Period	DEA condition satisfied *	TVE classification**	DEA model
University libraries	37	7	2000 - 07	Yes	Contemporaneous	BCC-O

Notes: -*: number of DMUs not less than two (three) times the number of variables; **: classification of (sample) observed subsets by Tulkens &Vanden Eeckaut (1995, p. 479-480).

We focus on Brazilian data collected from the centralized MIS maintained by the system of academic libraries pertaining to a traditional federal university in Rio de Janeiro. Our example is supported by a convenience sample of 37 units corresponding to more than 80% of the total population. Time periods refer to 2000 – 2007. Data relate to three inputs (*number of employees*, *physical area in square meters* and *number of volumes*) and four outputs (*number of visits*, *of loans*, *of registers* and *of consultations*).

Efficiency Analysis

The efficiency of productive units has been calculated by means of a Data Envelopment Analysis (DEA) deterministic frontier model as the solution of a linear programming problem (see references in Carvalho *et al*., 2012). Given the *a priori* restricted nature of public budgets, the output-oriented version was adopted. The solution of the appropriate linear programming problem provides numerical scores for each DMU and characterizes them with respect to efficiency status. For each inefficient DMU an operation plan is also provided that indicates (re)allocative targets for the DMU to reach efficiency. Finally scores will again be needed to compute the transitions between the two states along the time period for the whole set of libraries.

Markovian Analysis

As soon as a transition matrix is available, passage times and long run analysis are possible and will result from the computation of a fixed point for the transition matrix (Kemeny and Snell, 1972, p. 130-131). In order to get a transition matrix from empirical data it suffices to use the *transition count* (Wang and Huang, 2007, p. 1306; see also references in Carvalho *et al*., 2012) corresponding to the proportion of units in a given state and then count the transition between each pair of states in the period.

According to Kemeny and Snell (1972, p.131), when the number of time steps grows indefinitely one has

$$\lim (1/n)(P + P^2 + . . . + P^n) = [1 \ \ 1 \ \ 1 \quad 1]'\pi \qquad (3.2),$$

where n is the number of steps; $P^n = ((p_{ij}^{(n)}))$ is the nth power matrix of the one-step transition matrix P, whose (i ; j) element then represents the probability of transition from state i to state j after n steps; [1 1 ,…, 1]' is a column vector with all elements equal to 1, and π is precisely the fixed point, that is, a constant vector containing the long run equilibrium distribution between states whose components are nonnegative and sum to 1 (as any probability vector), and such that $\pi P = \pi$. The expression "long run equilibrium" is then adequate since in (3.2) π does not depend neither on time, nor on the initial state.

Since one has

$$\lim (1/n)(P^n + P^{n+1} + . . . + P^{2n}) = [1 \ \ 1 \ \ 1 \quad 1]'\pi. \qquad (3.3),$$

any power of the one-step transition matrix could be used to compute the fixed point π.

The first application of Markovian Analysis amounts to computing mean first passage time and mean recurrence time (Kemeny *et al*., 1964, p. 411-414) for any state of the system. A possible link between the mean first passage time and efficiency analysis stems from the fact that the time before the (mean) first passage into efficiency may suggest how urgent may be the changes indicated in the "operations plans" provided by efficiency analysis (corresponding to the second step in the proposed procedure).

The second application of Markovian Analysis is also related to the fixed point of P^n, since π directly provides the long run equilibrium of the system (Kemeny and Snell, 1972, p. 131). This equilibrium may be interpreted as the long run (percent) distribution of units between states, since system transitions between states are defined as counts of units' transitions.

4 Results: description and discussion

In this section findings are presented relating to the selected academic libraries. Comments follow the order of proposed steps – computed efficiency scores, operation plans and long run distribution.

First step – efficiency scores and library rankings

A sample profile for the 37 DMUs (see Table 1) is given in Table 2 for the last year of the period of study. Accordingly, the coefficients of variation imply that the libraries are quite different from one another on most attributes.

Table 2 – Sample profile for university libraries in 2007

Variables	Min	Max	Mean	Standard deviation	Coefficient of Variation
Employees	1	33	8,41	8,06	95,83%
Total area (m2)	37	6000	865,16	1400,03	161,82%
Volumes	872	277134	35228,92	53343,38	151,42%
Visits	108	137385	20974,68	33970,98	161,96%
Registrations	0	5603	1043,38	1115,40	106,90%
Loans	0	30191	5116,03	6578,68	128,59%
Consultations	0	66638	8091,62	12228,71	151,13%
Service mix (number)	5	13	9,54	1,87	20%

Computed efficiency scores appear in Table 3. Since every efficient DMU has a score equal to 1, the 8 libraries in that situation along 2000-2007 have been removed from Table 3. By the very definition of efficiency, there is no way to improve their productive performance: these DMUs present a quite robust performance and deserve attention no matter how “benchmark” is understood. Relatively inefficient DMUs receive a score less than 1. Note that some inefficient libraries never visited the efficient frontier and are even far away of it; in that sense they also deserve managerial attention. Note also that library number 5 has been efficient along the whole period except for one year. Should this situation be ascribed to measurement error? Does it mean a “true” although negligible loss in performance? In terms of management action all these signals must likely be accompanied by an individual follow-up.

Second step: optimal changes for each library along the period

Operation plans are summarized in Table 4 and deserve managerial attention since resource decreases may occur simultaneously with output increases, so that managers must keep alert and proactive as to take advantage from potential efficiency gains along time. Volume discards deserve special attention because some collections and some individual titles should not be altered.

Allocative changes such as those indicated in Table 4 may also serve to compare recommended paths against observed actions in a yearly basis, for each DMU, and to that extent may help evaluate individual performance.

Third step: first passages, mean recurrence and long run distribution

Given that we are working with contemporaneous reference sets (see Table 1), data for 2000-2007 provide an empirical version for the n-step transition as the seven factor product of the seven observed one-step matrices, say A = $P_1P_2 \quad P_6P_7$. The "seven factor product" approach simply amounts to envisage the long run as starting from the transitions occurring from the seventh year on and improve upon the "averaging" approach (CARVALHO *et al*., 2012).

Table 3 – Efficiency scores* and yearly averages : 2000 – 2007

DMU	SCORES 2000	SCORES 2001	SCORES 2002	SCORES 2003	SCORES 2004	SCORES 2005	SCORES 2006	SCORES 2007
1	1,000	0,841	1,000	1,000	0,605	0,811	0,680	1,000
2	0,571	1,000	1,000	1,000	0,965	0,943	1,000	1,000
3	0,305	0,936	0,845	0,661	0,542	0,846	0,775	0,574
4	0,989	0,960	0,769	0,783	0,829	1,000	1,000	1,000
5	1,000	1,000	1,000	1,000	1,000	0,947	1,000	1,000
6	1,000	0,696	0,742	0,494	0,584	0,757	0,548	0,650
7	1,000	0,731	0,870	0,452	0,353	0,127	0,466	0,624
8	0,941	1,000	0,471	0,559	0,782	0,650	0,626	1,000
10	0,620	0,895	0,712	0,974	0,619	0,740	1,000	0,679
11	0,528	0,660	1,000	0,779	0,727	1,000	0,847	0,646
12	0,404	0,590	0,287	1,000	1,000	1,000	1,000	1,000
17	1,000	1,000	0,627	1,000	1,000	1,000	0,336	0,370
18	0,604	0,815	0,696	1,000	1,000	1,000	0,807	1,000
19	1,000	1,000	1,000	1,000	1,000	0,959	1,000	0,921
20	0,600	1,000	0,867	0,779	0,743	0,498	0,543	0,560
21	0,401	0,302	0,396	0,109	0,138	0,371	0,145	0,115
22	1,000	1,000	0,507	0,654	0,337	1,000	0,842	0,121
24	0,391	0,501	0,492	0,387	0,395	0,931	0,319	0,320
25	0,733	0,690	0,840	0,329	0,307	0,482	0,640	0,506
26	0,838	1,000	0,467	0,683	0,236	0,562	0,384	0,863
27	0,334	0,412	0,410	0,407	0,358	0,223	0,496	0,241
28	0,892	0,574	1,000	1,000	1,000	1,000	1,000	0,945
30	1,000	0,442	1,000	0,555	0,972	1,000	1,000	0,820
31	0,071	0,064	0,055	0,143	0,185	0,020	0,010	0,017
32	0,450	0,781	0,928	0,873	0,870	1,000	1,000	1,000
34	0,562	1,000	1,000	1,000	1,000	1,000	1,000	1,000
35	1,000	0,793	1,000	0,665	0,757	0,354	1,000	1,000
36	0,107	0,202	0,196	0,172	0,113	0,353	0,401	0,381
37	0,359	1,000	1,000	1,000	0,892	1,000	1,000	1,000
Mean (n=37)	0,7486	0,8077	0,7886	0,7691	0,7381	0,7993	0,7801	0,7663
Pct effic.	45,96%	48,65%	48,65%	48,65%	40,54%	51,35%	54,05%	51,35%

Note. * - All libraries with scores equal to 1 for the whole period have been excluded.

From the fundamental matrix, the mean first passage time from "inefficiency" to "efficiency" is computed as approximately equal to 1 year and 10 months (KEMENY *et al*., 1964, p. 411). This means that if a given unit is inefficient today and if no managerial action is taken, then on average it will take 22 months for the unit to become efficient. This delay may be compared to the time required for any possible remedial measures to become effective, say revised budgeting or training.

Table 4 – Average operation plans : 2000 - 2007

Inputs	2000	2001	2002	2003	2004	2005	2006	2007
Employees (number)	- 1,44	- 1,15	- 0,76	- 1,29	- 0,93	- 1,18	- 0,61	- 0,81
Area (m2)	- 60,75	- 71,04 *	- 29,85	- 70,35	- 48,94	- 143,47	- 88,05	- 136,27
Volumes (number)	- 3064,48	- 3373,49	- 1880,71	- 4601,0	- 6447,08	- 651,77	- 4720,75	- 3153,65

Note * - this figure relates to a single library.

Considering (3.4), to obtain the (estimated) long run distribution of the system between the two states the fixed point equation $\pi A = \pi$ must be solved to give:

π_E (percent efficient) = 51, 5%; π_{NE} (percent inefficient) = 48,5% .

Note that π_E , the percent efficient, differs from the mean and the median percent efficient (48,7%) in the last line of Table 3. Remember that, in contrast to "short run averaging", products of transition matrices bring into play all the transitory visits to the two states along the time span.
The fixed point π in the equation $\pi A = \pi$ also provides directly the mean recurrence time (Kemeny *et al.*, 1964, p. 413) for the states of the system, that is, the mean time required before the system returns to a given state having started in that same state. The mean recurrence time is approximately equal to 2 years in both cases, so that the period of two years seems to be critical in the sense of monitoring the return of a state to itself. In the case of inefficiency it represents a sort of "safe mean time span" for managers to try to change the operating conditions facing inefficient units, Since the operation plans already point to "optimal changes" by unit, managers may evaluate for which units those changes would be feasible within (the next) two years. Note that on average an inefficient unit will return to inefficiency four months before it may reach efficiency for the first time, if no managerial action is taken.

5 Concluding comments

We proceeded in three steps. In the first two steps, typical DEA models provided quantitative indicators – namely, rankings and operation plans - that not only help evaluate library performance, but may also assist inefficient library units in their quest for efficiency.
In the third step we relied on Markov Chains for long run assessment. From a methodological viewpoint the model extends and improves upon previous work (CARVALHO *et al*., 2012). We first computed an aggregate measure of the distribution of the productive system (the "organization") between two states – efficient or inefficient. The other useful application of the Markovian approach provides better knowledge concerning the time delay required for efficiency to be attained for the first time when a prescribed operation plan happens to be adopted, as well as about the time during which an undesired (inefficient) situation will persist if that adoption is postponed. This timing aspect may help library managers in preparing their planning and control schedules and figures with a view toward the efficiency endeavour. For example, an inefficient unit will on average return to inefficiency four months before it attains

efficiency for the first time, so that managerial attention to such time lags may become critical .
Future research is likely to provide better theoretical as well as empirical information that will allow for a better assessment of the proposed model. In particular, since the long run is here depicted in a very simple way, the “short memory” assumption involved in Markovian approaches may appear inappropriate in many contexts. The adequate approach to this issue still requires more work.

References

Carvalho F.A., Jorge M.J., Jorge M.F., Russo M. and Sá N.O. (2012), “An optimizaton approach to change management in academic libraries with an empirical application in Rio de Janeiro, Brazil”, in Yildiz, A.K. (Ed.), *Management and Organisation in Information-Records Centers*, BETA Publ., Istanbul, Turkey, pp. 103-120.

Kemeny J.G., Mirkil H., Snell J.L. and Thompson G.L. (1964), *Finite Mathematical Structures*, 5th printing, Prentice-Hall Inc., Englewood Cliffs, NJ, USA.

______ and Snell J.L. (1972), *Mathematical Models in the Social Sciences*, The MIT Press, Cambridge, MA, USA.

Tulkens, H. and Vanden Eeckaut, P. (1995), “Non-parametric efficiency, progress and regress measures for panel data: methodological aspects”, *European Journal of Operational Research*, Vol 80 No. 3, pp. 474-499.

Wang M.H. and Huang T-H. (2007). “A study on the persistence of Farrell’s efficiency measure under a dynamic framework”, *European Journal of Operational Research*, Vol 180 No 3, pp. 1302-1316.