Libraries Research Publications

9-26-2002

Cost Efficiency in ARL Academic Libraries

E. Stewart Saunders
Purdue University, ssaunder@purdue.edu

COST EFFICIENCY IN ARL ACADEMIC LIBRARIES*

In the past decade the educational process has come under closer scrutiny from its stakeholders. Their major concern is how to assess the impact of instruction on students. The other side of the coin is how to contain costs in the process. It has long been a standing assumption that libraries are the least cost providers of information. This assumption is currently being challenged in the market place as alternative avenues to information are made available electronically. As budgets are squeezed, the long term outlook for research libraries may be an issue of cost. The purpose of this research is to inquire into the cost efficiency of academic research libraries.

Most of the literature on costs in libraries has focused on such issues as returns to scale, allocation of inputs, or unit costs for a single service. When these studies have used economic theory, the assumption has been that library managers minimize costs, that is, libraries are cost efficient institutions. This claim is made most often for public libraries (Deboer, 1992) (Hammond, 1999). In fact the argument for public libraries has some merit. To my knowledge such claims for research libraries have not been forthcoming.

Programmed budgeting or activity based costing offer one approach to the measurement of library cost efficiency. The line item budget, which is the standard method used in libraries, measures the cost of inputs. Programmed budgeting measures the cost of outputs. Robinson and Robinson (1994) use programmed budgeting to establish unit costs for services in five California public libraries. For instance, Library A

* An earlier version of this paper was read at the Second Library Research Seminar, University of Maryland, Nov. 3, 2001. I wish to thank Prof. Larry DeBoer of the Purdue University Dept. of Agricultural

had the lowest unit cost, $1.63, to circulate one book. If Library B with a unit cost of $2.34 for circulation wishes to be 100% efficient in that service, it must reduce unit cost to $1.63. What we have done here is to make Library A the "best practices" library for circulation and the standard against which all other libraries are measured. It is difficult, however, to convert efficiency ratings of individual services into a measure of total library efficiency. Each library delivers a different mix of services. Some may emphasize circulation, others interlibrary loan, or others reference. In order to arrive at a rating for a library's total cost efficiency it becomes necessary to weight each service. In the market place prices for outputs serve as this weighting mechanism, but in the public sector there are no prices for services. The issue then becomes, what makes a justifiable weight? Another approach to efficiency rating is the use of single factor productivity ratios. ARL provides these ratios as part of its statistical service. This again raises the issue of how to obtain a single efficiency rating for a library from multiple ratios without the use of a weighting system.

Frontier methodologies offer an alternative to programmed budgeting and productivity ratios for obtaining an overall measure of library cost efficiency. These methods look at all outputs collectively, weighting each output based on either a theoretical cost function or on the belief that a library should be compared to a "best practices" library which weights each service in the same way as the library being evaluated. The research described here measures the cost efficiency of 88 ARL academic libraries using frontier methods. The purpose of this paper is three-fold: (1) to expose library researchers to newer methodologies in cost analysis, (2) to determine whether

Economics for several helpful suggestions which improved the paper. I also wish to thank the Purdue Libraries for grant support to complete this project.

academic research libraries are cost efficient compared to other institutions or industries, and (3) to identify possible explanations where inefficiency is discovered.

Researchers have applied frontier methods to libraries, but the usual emphasis has been more on productive efficiency than on cost efficiency. These include Sharma, Leung, and Zane(1999), Worthington(1999), Vitaliano(1997 and 1998), Chen(1997), Shim and Kantor(1998), and Reichmann(2001). Most of this research appears in the economic and management literature rather than in the library science literature.

Basic Concepts

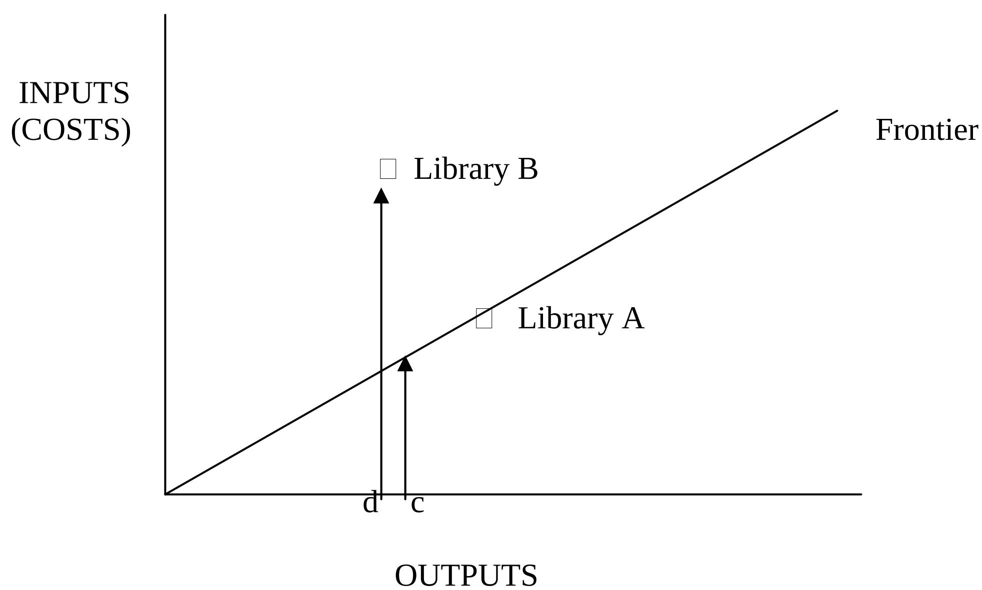
An economic definition of efficiency would state that a library will be 100% cost efficient if it is both technically efficient and allocatively efficient. A library is technically efficient if no increase in outputs is possible for a given amount of physical inputs. Physical inputs would be staff, materials, and facilities. A library is allocatively efficient if these physical inputs are made in the correct proportions. The correct proportion is attained when one physical input is substituted for another to obtain the lowest cost without reducing the level of output. The unit price of each input determines the lowest cost. This economic definition is often referred to as "Pareto efficiency." This is of course an ideal model based on the assumption that a manager has perfect and costless information about production and the prices of inputs. While this is never the case in the real world, modern information technology is allowing managers to approach this goal.

Cost efficiency can be conceptualized and measured in terms of a frontier. A cost frontier is a function which gives the minimum cost for any level of output. If a library is on the frontier, it will be 100% cost efficient. If it is not on the frontier, then its distance from the frontier tells the level of efficiency. In economics a cost frontier is defined by the cost function for a given industry. Another way to determine a frontier is to use the "best practices" concept. Those libraries with the least cost for a given level of output form the frontier and are the standard against which all other libraries are measured. A cost frontier works particularly well for libraries because it treats outputs such as circulation or reference questions as exogenous variables. This means that they are not under the control of management. Management controls costs, so it is the responsibility of management to establish whatever expenditures are necessary to delivery an output that is derived from external demand.

In the past 25 years econometricians have devise two quantitative procedures for using the frontier concept. Each involves a means of quantitatively defining the frontier and then measuring the distance from the frontier as a level of efficiency. One method is called data envelopment analysis (DEA). It puts cost and output data into a linear programming model which finds a frontier and distances of each library from that frontier. A second method is called stochastic frontier regression. It assumes a particular cost function and then uses cost and output data to fit a regression equation. The regression line becomes the frontier and all libraries are compared to this frontier for an efficiency rating.

Both methodologies have a common ancestor, an article written by M. J. Farrell (1957) entitled, "The Measurement of Productive Efficiency." Diagram 1 below illustrates the concept. The diagonal line is a minimum cost frontier. It shows the minimum cost for a given level of output. Library A is on the frontier and is therefore 100% efficient. Library B is above the frontier. In order to be 100% efficient Library B must lower its cost without reducing its output until it reaches the frontier. The efficiency rate for library B is measured as the ratio of line c to line d. Farrell's contribution is to demonstrate that this simple diagram can be applied to the economic definitions of production functions and cost functions where there may be multiple inputs or multiple outputs.

DIAGRAM 1



Having stated the possibilities offered by frontier technologies, it is necessary to state a caveat. Frontiers defined by stochastic regression or DEA will be Pareto efficient only if they meet certain conditions. These conditions are technical requirements built into theoretical frontiers, but frontiers fitted from data may not conform to these requirements. Thus a library on the frontier will be efficient by Farrell's definition but this does not guarantee that it will be efficient by the Pareto definition. For that reason we distinguish between Pareto efficiency and Farrell efficiency.

The Data

Data were acquired from *ARL Statistics 1998/99* and from *ARL Supplementary Statistics 1998/99*. At the time of the research the latest supplementary data available were for 1998/99. For that year there are 112 academic ARL libraries. Only libraries with complete data for all of the variables used were chosen. One library was rejected for what appear to be irregularities in its data. A total of 88 libraries were used in the analysis. Since these 88 libraries represent most of the academic libraries in ARL, no attempt is made to estimate population parameters.

Library outputs were defined in terms of patron use. Both Robinson and Robinson(1994) and Shim and Kantor(1998) also use patron use as a measure of output. The variables used are number of circulations, number of reference questions, number of bibliographic instruction sessions, number of interlibrary loan items borrowed and number loaned, and weekly hours of service. All of the outputs are direct measures of patron use except service hours. Service hours is a proxy variable for facility use. Many students use library facilities to study without ever borrowing a book or asking a reference question. Service hours should capture this use indirectly. The number of volumes added in a year or the number of serial subscriptions were not used as an output measure. We may think of these as intermediate outputs in a vertically integrated industry. Acquisition does not equal use. As of yet, ARL data does not include use of electronic sources. Inclusion of this data would make the analysis more complete, but we must await the decision of ARL on its inclusion.

Library inputs were measured as expenditures on materials, salaries, and binding. This included all of the expenditure data reported in the ARL statistics except a category labeled "Other operating expenditures." This latter data was excluded since it is not consistent across libraries. Library cost is the sum of expenditures on materials, salaries, and binding. By this measure costs include material and labor cost but not capital costs.

Stochastic Frontier Regression Analysis

Stochastic frontier regression uses a specific cost function as the determined part of a regression equation. Since the cost function defines a frontier, deviations from the determined part may be considered as measures of inefficiency. The deviations are of course in the error term of the regression equation. The problem is, however, that a least squares fit gives both negative and positive errors with a mean of zero. In order for the error to be a measure of inefficiency it must be positive because these represent over expenditures. Stochastic frontier regression solves this problem be dividing the error term into two parts. One part is pure noise with a mean of zero. The second part has only positive errors and represents excessive costs, i.e., a measure of inefficiency.

$$\text{Error} = \text{Noise} + \text{Excessive costs}$$

If we assume a distribution for each part of the error, then the regression equation can be fitted using maximum likelihood methods. One may for instance assume that the pure noise portion of the error is normally distributed and that the excessive cost portion is distributed like the positive half of a normal distribution. When the two distributions are added together we have a distribution which has a positive skew. The more the skew, the greater is the amount of inefficiency. See Kumbhakar and Lovell (2000) for a derivation and explanation of the density functions needed to fit this equation.

For this research project it was assumed that the cost function for ARL libraries is a Cobb-Douglas cost function. Other researchers (Vitaliano, 1997) (Vitaliano and Toren, 1994) have found the Cobb-Douglas preferable to the Translog function. Under simplifying assumptions the latter cost function reduces to the Cobb-Douglas. The following variables were used to fit a regression equation.

Y = Total Library Cost

X_1 = Constant

X_2 = Number of Circulations

X_3 = Number of Reference Questions

X_4 = Number of ILL loans

X_5 = Number of ILL borrows

X_6 = Number of Bibliographic Instruction Sessions

X_7 = Number of Weekly Hours of Service

X_8 = Average Professional Salary

X_9 = Average Non-Professional Salary

A cost function would normally also include a price for capital and a price for materials. The capital rate is not included among the variables because the expenditure data used does not include capital costs. By not including a price for materials, it is assumed that each library pays the same price for the same item, i.e., the subscription price for the

Journal of Organic Chemistry will be the same to one library as to another. In so far as this assumption is not true, the library paying the lower price will erroneously be considered more cost efficient. It is recognized that some libraries do receive price breaks that other libraries do not receive, but it is hoped that these differentials will be small enough that the effect on efficiency ratings will be minimal.

The regression equation for the Cobb-Douglas cost function is shown below. As stated above, all terms on the right hand side of the equation except the error term make up the Cobb-Douglas cost function and thereby identify a deterministic frontier which applies to all libraries. Because noise is inherent in any data fitting technique, the cost frontier for each individual library is composed of the Cobb-Douglas cost function plus the noise term. Deviations from the individual cost frontiers are measured as the excessive cost term.

$$\mathrm{Ln}(Y/X_9) = X_1 + b_2 \,\mathrm{Ln}\, X_2 + b_3 \,\mathrm{Ln}\, X_3 + b_4 \,\mathrm{Ln}\, X_4 + b_5 \,\mathrm{Ln}\, X + b_6 \,\mathrm{Ln} X_6 + b_7 \,\mathrm{Ln} X_7 + b_8 \,\mathrm{Ln}(X_8 / X_9) + \mathrm{Error}$$

$$\mathrm{Error} = \mathrm{Noise} + \mathrm{Excessive\ costs}$$

X_9 is used as a divisor so that the cost frontier will be linearly homogenous with respect to prices. This equation was fitted to the data for 88 ARL academic libraries using LIMPED 7.0. The noise was assumed to be normally distributed and the excessive costs half-normally distributed.

Data Envelopment Analysis

In order to determine the cost efficiency of a library, data envelopment analysis compares that library to a "best practices" library, i.e., one on the cost frontier. DEA does this in two linear programming steps. In the first step DEA identifies a library on the frontier that has the identical output as the library being evaluated for efficiency. Thus each library would have the same number of circulations, the same number of reference questions, etc. Only the costs between the two libraries would differ. Since the "best practices" library would have the lower cost, it is possible to determine the cost efficiency of the library being evaluated by a simple comparison of costs. The formula is: Expenditures of "best practice" library divided by Expenditures of the library being evaluated for efficiency. This gives the Farrell cost efficiency. If the libraries had only one output, e.g. circulation, then the whole procedure could be done with a simple plot like Diagram 1 above. For the ARL libraries six output measures were used. Since it is unlikely that any two libraries have identical levels on all outputs, DEA creates a virtual library for comparison purposes. This virtual library is a synthesis of several real libraries on the frontier and is itself on the frontier. In creating a virtual library DEA defines one that has all the same level of outputs as the one being rated for efficiency. Then the cost of the library under investigation is compared to the hypothetical cost of the virtual library on the frontier, and a Farrell efficiency rate is computed.

In the second step DEA attempts to identify a second "best practices" library which has the same expenditures as the first "best practices" library but has a greater level of output. If DEA cannot find one, then the first "best practices" library is both Farrell efficient and Pareto Efficient. If it does find one, then the second "best practices" library becomes the standard for measuring Pareto efficiency. A complete understanding of how

DEA works requires a knowledge of optimization techniques and linear algebra. For an intermediate introduction to DEA see Cooper, Seiford, and Tone (2000).

For DEA the following variables were used:

OUTPUTS	INPUT
1. Number of Circulations	1. Normalized Costs
2. Number of Reference Questions	
3. Number of Bibliographic Instruction Sessions	
4. Number of ILL Loans	
5. Number of ILL borrows	
6. Weekly Hours of Service	

Library costs were normalized by setting expenditures for professional and non-professional staff to a common annual salary for each group. The common salary used is the average of all the salaries in 88 libraries, for each group. In cost theory different wages for different libraries are treated as exogenous variables and hence are not a factor in determining cost efficiency. The Cobb-Douglas cost function (see above) controls for differential wage rates by including wage rates in the cost function. In DEA control is attained by using normalized expenditures for staff. The formula for normalized costs is:

Normalized costs = normalized staff expenditures + materials cost + binding costs

The analysis using DEA was carried out using the Excel Solver as a linear optimizer. The optimization formula below folds both steps of DEA into one set of equations.

Minimize:

$$Z = R - e\sum_{k} s_j^{+} - e s^{-}$$

(The value of **R** gives the cost efficiency. **S** is a slack value put there to balance the equation. If the values of all the **S**s are zero, then **R** denotes both Farrell and Pareto efficiency; if they are not zero, then **R** denotes only the Farrell efficiency. **e** has a value of 1E-10.)

Constrained by:

$$R\ I_o = \sum_{h} L_i I_i + s^{-}$$

(**Io** is the input value, i.e., the normalized cost, for the library being evaluated for efficiency. **Ii** is the input value or cost for library **i**. Notice that it is summed for all 88 libraries. **Li** is the weight for library **i.** The linear program assigns the weight. If library **i** is not on the frontier, then **Li** must be zero.)

$$O_o^k = \sum_{h} L_i O_i^k - s_k^{+}$$

(**Oo** is one of the outputs for the library being evaluated for efficiency. **Oi** is one of the outputs for library **i**. Here again it is summed for the 88 libraries. **Li** is the weight for library **i**. Since there are six outputs, circulation, reference, ILL, etc., there must be six of these equations. That is the reason for the superscript **k**. **k** assumes the values 1 through 6.)

$$all\, L >= 0$$

$$all\; s >= 0$$

$$R >= 0$$

$$\sum L = 1$$

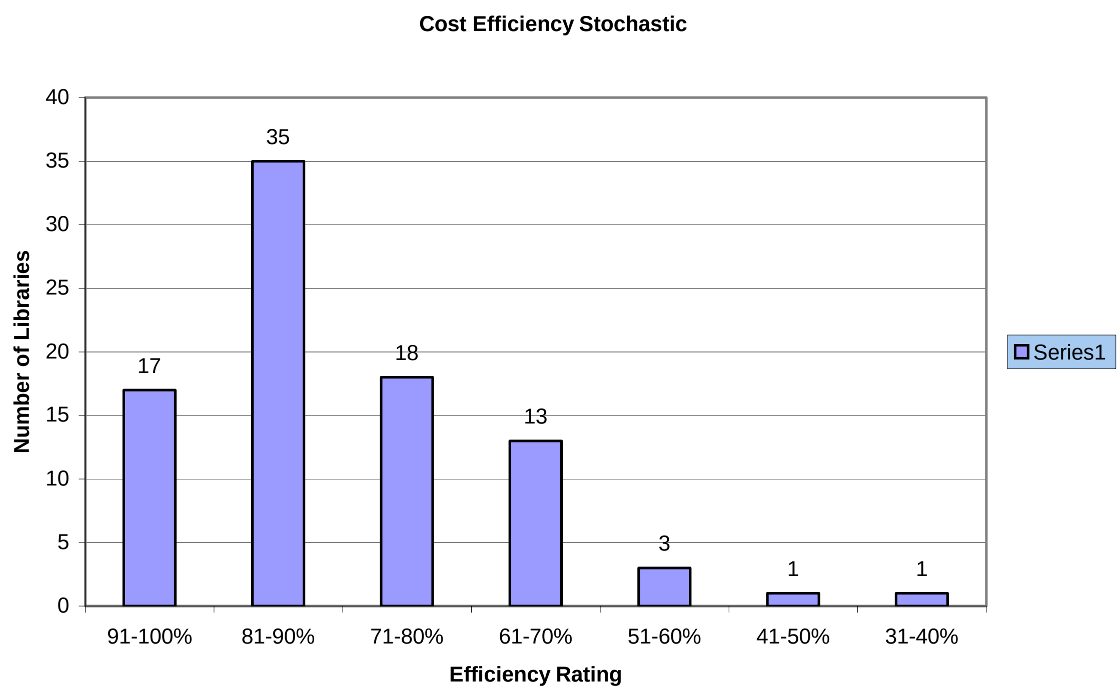
(These equations state that **L**, **S**, and **R** are not allowed to assume negative values when the optimization program is run. Variable returns to scale require that all the values of **L** sum to 1.) The linear program above is applied 88 times, once for each library being evaluated.

The Results

A. Stochastic Frontier Regression

A stochastic cost frontier will only give a fit if there is a right skew to the errors. The program fitted the data; the fit gave an R^2 of 72%. A distribution of the libraries by efficiency rate is shown below in Chart 1.

CHART 1

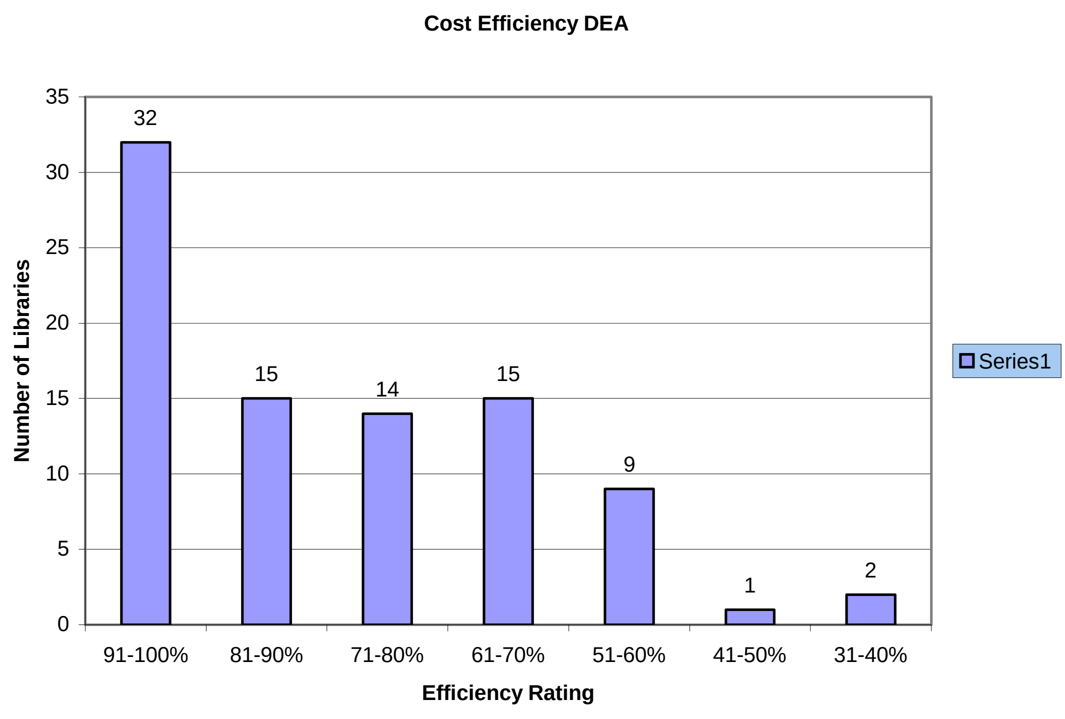


Most libraries fell in the 81% to 100% range. What confidence can we place in the results, or must there be reservations? One cannot be certain that a Cobb-Douglas cost function best describes the technical processes of a library or that excess costs will be half normally distributed. These are assumptions made in order to obtain a fit. Cobb-Douglas does meet the conditions required by economic theory for Pareto efficiency, although this benefit may be compromised when multiple outputs are present. On the plus side, the presence of a noise term allows us to filter out errors of measurement or outcomes not under the control of management. When looking at the individual library these measures should be compared with the DEA efficiency measures.

B. Data Envelopment Analysis

A distribution of the libraries by efficiency rate using the DEA model is shown below in Chart 2.

CHART 2



As with stochastic frontier regression most libraries are between 81% and 100% efficient. Again, what confidence can we place in these results? One advantage of data envelopment analysis is that no particular cost function is assumed, nor is a particular distribution of excess costs assumed. Everything is derived entirely from the data. Each library is fitted individually to the data. This means that each library has its own cost function reflecting its own goals. One can think of DEA as a weighted sum of single factor productivity ratios. Unlike stochastic frontier regression, however, DEA does not have a noise term. There is no way to allow for errors of measurement or outcomes outside management's control.

Unfortunately a fit using DEA does not always meet the conditions needed for Pareto efficiency. This means that the efficiency score will be an accurate Farrell measure of efficiency but not necessarily an accurate Pareto measure. To illustrate this look at the DEA results below in Tabel 1 for three libraries.

TABLE 1

FERRALL AND PARETO EFFICIENCY USING DEA

Library	Actual expenditure	Frontier expenditure	Ferrall Efficiency	Pareto Efficiency
A	$9,017,047	$9,017,047	100%	100%
B	$16,643,387	$11,046,611	66.4%	66.4%
C	$13,563,280	$12,210,252	90%	90% plus increase: a. reference by 18, 387 questions b. hours by 24

Library A is on the frontier since its actual expenditure equals the expenditure of a frontier library. It is 100% cost efficient by both a Ferrall measure and a Pareto measure. DEA has found a first "best practices" library on the frontier with which to compare

Library B. To attain this position on the frontier, Library B would have to reduce its costs to 66.4% of $16,643,387, i.e., to $11,046,252. DEA cannot find a second "best practices" library on the frontier, so Library B cannot increase any output without also increasing costs. Thus this frontier position is both Farrell efficient and Pareto efficient, and we can use the same percentage to state the Farrell and Pareto efficiency of Library B. DEA has found a first "best practices" library on the frontier with which to compare Library C. If Library C reduces its costs to 90% of $13,536,280, it will be 100% Farrell efficient. But DEA also finds a second "best practices" library on the frontier where Library C must also increase the number of reference questions by 18,387 and extend hours by 24 and hold its costs to $12,210,252. So 90% is a measure of Farrell efficiency but not of Pareto efficiency. For Library C there is no single number to measure Pareto efficiency.

What does this mean for the individual library? Farrell efficiency will always be a single number and will be equal to or higher than a Pareto rating. All of the ratings used in the DEA analysis are Farrell ratings. In some instances the Farrell rating will also be the Pareto rating; in other instances it will not. This means that each library is put in its best light. If it were possible to have a true Pareto rating as a single number, then many of these libraries would have lower ratings.

C. A Comparison of Results

Table 2 gives the cost efficiency using both methods for a sample of 20 libraries.

TABLE 2

A COMPARISON OF LIBRARY EFFICIENCIES

FOR 20 LIBRARIES USING

DEA AND STOCHASTIC FRONTIERS

LIBRARY	DEA	STOCHASTIC FRONTIER
1	92.9%	78.6%
2	100.0%	87.7%
3	63.5%	72.5%
4	100.0%	86.8%
5	78.5%	81.8%
6	90.4%	84.1%
7	100.0%	87.1%
8	73.5%	70.8%
9	60.1%	72.6%
10	76.8%	90.9%
11	59.2%	61.2%
12	97.5%	91.3%
13	73.5%	87.8%
14	100.0%	85.4%
15	92.5%	92.6%
16	100.0%	92.4%
17	36.7%	48.1%
18	75.5%	85.2%
19	50.6%	68.3%
20	79.2%	85.7%

The correlation of efficiency measures for the two methods is .861. Table 3 shows the average cost efficiency and the number of libraries on the frontier.

TABLE 3

AVERAGE EFFICIENCY AND FRONTIER POSITIONS

BY METHOD USED

LIBRARY	DEA	STOCHASTIC FRONTIER
AVERAGE LIBRARY EFFICIENCY	80.6%	80.5%
LIBRARIES ON THE FRONTIER	23	0

Interpreting the Results

The reader may ask, how can one really know the efficiency rating when each method gives a different result? Obviously different assumptions give different results. As one reads across the rows of Table 2, one sees that some libraries have very similar ratings by both methods while other libraries may differ by as much as 15% points. But even with a 15% spread in ratings, it is still possible to place libraries into high, middle, and low categories based on the two efficiency measures.

How sensitive is the efficiency measure to missing outputs? The all important use of electronic resources is lacking. Some libraries have archival and preservation goals. For them it is important to preserve the written record whether it is used or not. Yet other research libraries have begun to produce digital texts, an output outside the normal functions of a library. To obtain an idea of how important a missing output is to the measure of efficiency, the analysis can be rerun with one output purposefully removed. This gives a measure of how sensitive the methods are to missing variables. The stochastic frontier model was rerun six times with one of the outputs remove for each iteration. Stochastic frontier regression appears to be moderately sensitive to a missing output. The average effect of removing service hours, reference, and circulation, one at a time, was 6.8% to 8.3% on each library's measure of efficiency. The other three outputs had little effect. A detailed examination of the effect of dropping circulation from the stochastic frontier regression equation shows that while most libraries had little change in efficiency ratings, eleven libraries lost 15% or more in efficiency. For comparison purposes, a DEA analysis was also carried out with the circulation output missing. Eight libraries lost 15% or more in efficiency; all eight were among the eleven libraries effected using stochastic frontier regression. These results suggest that the inclusion of output data on use of electronic resources, or some other important output, could produce a moderate shift in efficiency for a number of ARL libraries.

What are the contributing factors to inefficiency in libraries? The answer to that question must be specific to a given library. As a group, however, it is possible to use statistical analysis to find variables which predict inefficiency. The variables may not only be predictive, but in some cases causal. A set of likely variables were chosen for analysis with the efficiency ratings from the DEA analysis. These variables were correlated with these efficiency ratings and regressed on these efficiency ratings. The results are given in Table 4 below.

TABLE 4

ANALYSIS OF THE SOURCES OF INEFFICIENCY

VARIABLES	Pearson Correlation with DEA Efficiency Rating	Standardized Coefficients from Regression on DEA Efficiency Ratings
Private University	- .394	- .204
Number of Current Serials	- .561	- .535
Volume Holdings	- .519	.441
Number of Students	.000	.298
Number of Faculty	- .233	.165
Ratio of Professional to Nonprofessional Staff	.002	- .041
Number of Library Staff (excluding student staff)	- .558	- .791
		N = 87 R^2 = .555

Based on the negative standardized coefficients, the variables which predict inefficiency are size of library staff, number of current serials, and private governance. While volume holdings have a negative correlation, their effect on efficiency is not negative when other variables are held constant. In a way this is logical since volume holdings do not represent a high annual cost in the same way that current serials do. Somewhat surprisingly, the ratio of professional to nonprofessional staff is not a factor in predicting efficiency. These results might lead one to conclude that larger libraries, i.e., ones with a larger number of serial subscriptions and a larger staff, are penalized in terms of efficiency ratings. This is not necessarily so. One library with a normalized expenditure of $29,000,000 was 100% Ferrall and Pareto efficient using DEA analysis.

As a group ARL libraries rate well in terms of cost efficiency. Using stochastic frontier regression, Vitaliano (1997) found public libraries in New York state to be 80.6% cost efficient; Vitaliano and Toren (1994) found New York state nursing homes to be 77.5% efficient; Ferrier and Lovell (1990) estimated commercial banking to be 79.3% efficient. (These ratings were reported as inefficiency scores. It was necessary to use their reciprocals in order to state them as efficiency scores). Using data envelopment analysis, Shim and Kantor (1998) found ARL libraries to be 96% efficient. Vitaliano (1998), using the data from his 1997 study of New York state public libraries, found a mean efficiency of 67% by using DEA. Both the Shim and Kantor study and the Vitaliano study used physical quantities for inputs rather than the expenditure data used in this study. This makes a direct comparison with this research difficult. By any of these comparisons the cost efficiency of ARL academic libraries is in the normal range.

Scale Efficiency

Cost efficiency is defined as a combination of technical and allocative efficiency. These are the efficiencies that can be attained by a good manager by controlling the technical processes and allocating inputs. Once these efficiencies are attained, however, further efficiencies can be gained through an appropriate scale of production. A library director cannot determine a scale of production since this is set by the size of the academic institution. It is interesting nonetheless to see where these economies may lie.

The variable returns to scale assumptions used with DEA allows us to determine the returns to scale for the libraries that are on the frontier(Cooper, Seiford, and Tone, 2000). Of the 23 libraries on the frontier 10 had increasing returns to scale. Their normalized expenditures ran from $7,800,000 to $10,000,000. Nine libraries had constant returns to scale with normalized expenditures ranging from $9,750,000 to $22,250,000. Four libraries had decreasing returns to scale with normalized expenditures ranging from $18,600,000 to $29,000,000. This runs counter to an earlier study which found that large research libraries have increasing returns to scale. (Cooper, 1984) Since constant returns to scale will have the lowest average unit costs for outputs, those libraries with budgets between $10,000,000 and $20,000,000 (normalized values) operate at the most efficient scale.

Conclusions

(1) Frontier methods give us the means to measure the overall cost efficiency of a library in a way that accounting and productivity ratio methods do not. Unfortunately neither stochastic regression nor data envelopment analysis is perfect. Each has strong and weak features. The strong feature of stochastic frontier regression is that it contains a noise element. A large body of data like the ARL data is certain to contain errors of reporting. These reporting errors can be filtered out as part of the noise in stochastic frontier regression. The strong feature of DEA is the lack of model assumptions. There are no assumptions about the form of the cost function nor the distribution of the over expenditures. Unfortunately, DEA contains no noise element. The perfect tool for cost analysis would be a DEA model with a noise feature. At this time so such tool exists. This should not, however, prevent us from using the two methods described here. While they give different efficiency rates for the same library, their rates are similar enough to give us a general idea of the cost efficiency of a library.

(2) Academic research libraries as a group are cost efficient. The average efficiency is around 80%. This compares very well with research results from both the public and private sectors of the economy.

(3) Given the discrepancies in our two methods and given that there is no data for electronic use, the most we can say about an individual library's efficiency is to classify it as high, middle, or low. Given a broad classification like this, however, some libraries are still less cost efficient than others. The DEA methodology used here allows for variable returns to scale in order that larger libraries with declining returns to scale will not be penalized in terms of efficiency for their size. One very large library was in fact 100% cost efficient. Even so, the variables (see regression coefficients of independent variables on DEA efficiency ratings) that explain, in part, lower efficiency are the number of current serials and the number of library staff, i.e., it can be traced to the size of the library.

There is no intent in this research to single out specific libraries for inefficiency without knowing the circumstances. The definition of efficiency used in this research is the degree to which libraries meet the information needs of its users while containing costs. Not all libraries will define there mission in terms of meeting the information needs of users, nor should they. Some libraries need to be inefficient by this standard; some libraries must assume leadership in preservation, building collections for no immediate use, or digitizing texts. Some libraries have mandates from their parent institutions in this regard. On the other hand, should a library have a low efficiency rating and it is pursuing none of these "cultural heritage" goals, then it needs to examine its procedures and organization for redundancy.

Bibliography

Chen, T. Y. 1997. "A Measurement of the Resource Utilization Efficiency of University Libraries," **International Journal of Production Economics** 53(1) 71-80.

Cooper, M. D. 1984. "Economies of Scale in Large Academic Libraries," **Library and Information Science Research** 6: 321-33.

Cooper, W. W., Seiford, L. M., and Tone, K. 2000. **Data Envelopment Analysis; a Comprehensive Text with Models, Applications, References and DEA-Solver Software**. Boston: Kluwer Academic Publishers.

Deboer, L. 1992. "Economies of Scale and Input Substitution in Public Libraries," **Journal of Urban Economics** 32: 257-68.

Ferrier, G. D., and Lovell, C. A. K. 1990. "Measuring cost efficiency in banking," **Journal of Econometrics** 46(1/2):229-45.

Hammond, C. J. 1999. "The technology of library service provision: a cost function analysis of public library systems in the United Kingdom," **Information Economics and Policy** 11:271-95.

Kumbhakar, S. C. and Knox Lovell, C. A. 2000. **Stochastic Frontier Analysis**. Cambridge: Cambridge University Press.

Reichmann, G. 2001. **Universitatsbibliotheken im Vergleich; eine Internationale Querschnittsuntersuchung**. Wiesbaden: Deutscher Universitats Verlag.

Robinson, B. M. and Robinson, S. 1994. "Strategic planning and program budgeting for libraries," **Library Trends** 42(3) 420-47.

Sharma, K. R., Leung, P., and Zane L. 1999. "Performance Measurement of Hawaii State Public Libraries: An Application of Data Envelopment Analysis," **Agricultural and Resource Economics Review** 28(2) 190-8.

Shim, W. and Kantor, P. B. 1998. "A novel approach to the evaluation of academic research libraries," **Proceedings of the 61st ASIS Annual Meeting** 35: 400-410.

Vitaliano, D. F. 1997. "X-Inefficiency in the public sector: the case of Libraries," **Public Finance Review** 25(6): 629-43.

Vitaliano, D. F. 1998. "Assessing Public Library Efficiency Using Data Envelopment Analysis," **Annals of Public and Cooperative Economics** 69(1) 107-22.

Vitaliano, D. F., and Toren, M. 1994. "Cost efficiency in nursing homes: a stochastic frontier approach," **Journal of Health Economics** 13:281-300.

Worthington, A. 1999. "Performance indicators and Efficiency Measurement in Public Libraries," **Australian Economic Review** 32(1) 31-42.