The Cost Function and Scale Economies in Academic Research Libraries

LEWIS G. LIU

ABSTRACT

THIS EMPIRICAL RESEARCH EXAMINED scale economies of academic research libraries that belong to the Association of Research Libraries and developed a total cost function for estimating economies of scale. The author argues that libraries in general, and academic research libraries in particular, are information provision organizations that provide multiproducts and multiservices and points out that some previous studies that used the production function have limitations due to the fact that the production function only permits a single output variable. This investigation incorporated a wide range of collections and service output variables into the total cost function. The regression results show that the adjusted R square of the cost function model is 0.8 and that the coefficients of three very important output variables (volumes held, serials, and group presentations) are statistically significant at high confidence levels. The final findings of this research show that the function coefficient is 0.93, indicating that slight economies of scale exist in academic research libraries.

INTRODUCTION

Libraries are important economic entities in modern society. At present there are 9,046 public libraries, 3,685 academic libraries, 98,169 school libraries, 9,763 special libraries, 1,376 government libraries, and 335 armed forces libraries with a total of over 122,300 libraries in the United States.[1] Improving efficiency and preventing misallocation of resources in libraries are as important as in other economic sectors of the national economy. One way to evaluate an organization's efficiency is to examine whether scale economies exist in the organization. The concept of scale economies is

Lewis G. Liu, City University of New York at Baruch College, 151 E. 25th Street, Box H-0520, New York, NY 10010
LIBRARY TRENDS, Vol. 51, No. 3, Winter 2003, pp. 293–311

rooted in economic studies of manufacturing. Massive production makes it possible for a firm to increase output and in the meantime reduce average cost to the extent where the firm needs to hire more people, purchase more equipment, and rent more facilities. Economies of scale exist because of many factors. New technologies and specialization of knowledge enable staff to work more efficiently. Purchasing a large quantity of supplies at a discount price can also result in substantial savings. Diseconomies of scale exist when output increases and average cost increases as well. Unskilled labor, outdated technologies, and poor management decisions can contribute to diseconomies of scale. Constant returns to scale occur when neither economies of scale nor diseconomies of scale exist. In many cases, researchers look at the relationship between marginal cost and average cost to determine whether or not excess capacity exists. If marginal cost (cost for producing an additional unit of output) is less than average cost, the firm is still better off by producing additional output.

Economic researchers have long been concerned with improving efficiencies of firms. Econometric models, such as production function, total cost function, and average cost function have been developed to measure economies of scale and to improve efficiencies of firms as well as government agencies and nonprofit organizations. Although many of the early research efforts focused on manufacturing and industries, such as railroads, metals and machinery, aircraft, gas, coal, telephone industries, and so on (Mansfield, 1997), later research began to study government agencies (Bauer, 2000), and educational institutions (Cohn, Rhine, & Santos, 1989; Koshal & Koshal, 2000; Chakraborty, Biswas, & Lewis, 2000). In the past, a few studies were conducted to examine economies of scale of libraries. Much of the previous research focused on public libraries. A handful of studies dealt with scale economies in academic libraries. Researchers wanted to know whether library sizes make a difference in terms of cost savings. They tested the assumption that larger-sized libraries tend to operate more efficiently than smaller ones.

Studies on scale economies have strong implications for public policymaking. The evidence that shows cost savings as a result of economies of scale can convince policymakers to make decisions for consolidating smaller-sized libraries (Baumol et al., 1969; DeBoer, 1992), although in many cases convenient service to local communities can be a more important factor than economic consideration. This investigation focused on academic research libraries that belong to the Association of Research Libraries. The purpose of this investigation was to examine whether economies of scale exist in academic research libraries. If economies of scale do exist, academic research libraries can, in the long run, provide better quality information services as well as increase the range of information services for faculty and students of their institutions at relatively less costs.

The missions and goals of academic research libraries are different from those of general academic libraries and public libraries (Liu, 2000). For example, academic research libraries have a stronger research orientation than general academic libraries that focus on teaching. To support their universities' mission for research, academic research libraries develop comprehensive collections that include a large number of research journals. Research journals consume a large portion of a research library's budget. Compared with academic research libraries, general academic libraries have only a limited number of research journals, and general public libraries have virtually no research journals at all. This and other differences between academic research libraries and general academic libraries and public libraries affect output factors in a cost function and warrant a new investigation.

Literature Review

The previous empirical research produced mixed results. Some researchers found evidence of increasing returns to scale. Others found evidence of decreasing returns to scale. Still others found evidence of constant returns to scale. The research on scale economies of libraries may be classified into three categories: studies that examined scale economies using the Cobb-Douglas production function; studies that looked at scale economies as well as elasticities of input substitution using the translog cost function; and studies that explored scale economies using general cost functions.

Early Studies

The cost study on libraries commissioned by the National Advisory Commission and conducted by William J. Baumol et al. in 1969 and another study by Stanley W. Black in the same year are probably the earliest studies on costs and economies of scale of libraries. Baumol et al. studied various types of libraries and examined total expenditures on staff and library materials, volumes owned and circulated, and growth rates. The report provided a library cost trend analysis and showed that library cost per capita and circulation unit cost decreased as sizes of public libraries increased, indicating economies of scale, but it also pointed out that savings were not expected to be substantial (p. 224). Baumol and Matityahu Marcus later studied costs of academic libraries, which led to the publication of a book in 1973. No production function or cost functions were dealt with in their book.

Studies Using the Cobb-Douglas Production Function

Stanley W. Black used the Cobb-Douglas production function for public libraries and treated circulation as the sole output, and labor and bookstock as inputs. The coefficients of the two observed variables were not statistically significant due to a high degree of multicollearity between them. As a result, he assumed that returns to scale were constant and was able to esti-

mate labor and bookstock elasticities, which were 0.833 and 0.167, respectively. Black's study seemed to have a far-reaching influence on the later studies since the methodologies used by the later studies tended to resemble his.

Haynes C. Goddard studied more than one hundred public libraries in Indiana using the Cobb-Douglas production function (1973). In his study, circulation was treated as output and bookstock, labor, materials, and capital were treated as inputs. Labor was measured based on circulation staff hours, and materials were measured in terms of their values by annual expenditures on them. Capital was measured by furniture associated with library services such as tables, chairs, card catalog, and so forth. He found that the function coefficient was 1.076, indicating slight increasing returns to scale. In addition, from a subsample, he estimated that the marginal costs of circulation were lower than average costs (p. 200), an indication of excess capacity in the libraries.

Robert M. Hayes used the Cobb-Douglas production function to study both public and academic libraries (Hayes 1979; 1981; Hayes & Borko, 1983). In the 1979 paper he studied optimal use of labor and capital by applying the Cobb-Douglas production function to large public libraries in California as well as in Illinois, Ohio, Wisconsin, and Missouri. He regressed circulation, a measure of output, on capital measured by the size of collection, and staff needed to maintain the facilities and the collection; and on labor measured by service staff. About 60 percent of the total variance in the dependent variable was accounted for by the Cobb-Douglas production function. Although Hayes used a single output measure, he recognized that other output factors, such as reference service, could also be considered. However, he pointed out that reference statistics available at that time might not be consistent and reliable. In his 1981 study on the use of library collections as measured by circulation and in-house use, he concluded that circulation did not adequately measure the usage of library collections. In 1983, he and Harold Borko published an article examining the relationship between library collections and faculty productivity using the Cobb-Douglas production function. The findings showed that library collections contribute significantly to faculty productivity as measured by faculty publications.

Studies Using the Translog Cost Function

Like Black (1969), Larry DeBoer and Christopher J. Hammond examined not only scale economies but also substitution elasticities of inputs (DeBoer, 1992; Hammond, 1999). However, they used a translog cost function instead of a production function. The translog cost function can be used to deal with issues of the demand for inputs of libraries as well as scale economies. This method is flexible in approximating production technologies in terms of input substitution elasticities. DeBoer used the translog cost function to examine economies of scale and input substitution elasticities of 194 Indiana public libraries. Book circulation was used as the output measure. Total cost was treated as the dependent variable. The translog cost function

required three input price variables for three inputs (labor, books, and supplies/services). The labor price was calculated by total labor expenditures divided by total number of FTE employees. He constructed a price index for supplies/services based on wages and prices in each district county. The relative book prices were used in the estimated equation. His findings showed that economies of scale existed for small libraries and constant returns to scale existed for large libraries. He also found that all production inputs were substitutes. Higher labor costs caused increases in purchases of supplies and services as well as books. Higher book prices led to an increase in purchases of supplies and services. But supplies and services were more responsive to higher labor prices than to book prices. Higher prices of supplies and services resulted in increases in purchases of both labor and books.

In a recent analysis of the cost function for UK public libraries, Christopher J. Hammond also used the translog cost function to examine scale economies and input substitution elasticities. He found that there existed increasing returns to scale and diseconomies of scope. Hammond also concluded that all production inputs were inelastic substitutes. What differentiated this study from some of the previous studies was that Hammond recognized the multiproduct nature of libraries, and used annual bookstock, audio-visual materials, and number of inquiries as outputs in the cost function.

Studies Using General Cost Functions

Not all researchers were interested in studying substitution elasticities of inputs. If the issue of the demand for inputs is not the research concern, a general cost function is sufficient for studying scale economies. The research conducted by Kathleen Foley Feldstein, Michael D. Cooper, and Paul Kantor used the general cost function approach (Feldstein, 1976; Cooper, 1979, 1983; Kantor, 1981a, 1981b). A general cost function can be logarithmic or nonlogarithmic. Economic researchers often transform a cost function into a common log form or a natural log form for the convenience of calculating the function coefficient or for the convenience of developing a model that is a better fit of a data sample.

Feldstein examined scale economies of public libraries using the national data and developed various cost functions. Since she used circulation as a single output measure, she was able to measure marginal costs from the total cost function, and average cost function using total costs divided by circulation. She found that although library systems had small diseconomies of scale, some individual libraries had economies of scale (p. 87).

Cooper examined whether or not there existed economies of scale in public libraries as well as academic libraries. Cooper's 1979 study collected data from public libraries in California. He recognized the fact that libraries provide multiproducts and services and regressed total expenditure on a number of output variables including volumes added, volumes borrowed, volumes lent, reference transactions, and circulations. He tested a number of alternative models, linear and nonlinear, logarithmic and non-

logarithmic and concluded that the log-linear model was the best fit of the data. He found that the function coefficient was slightly larger than 1 and argued that it should be interpreted as constant returns to scale. Cooper's 1983 study on academic libraries was similar to his 1979 study except that he added one more output variable: library hours opened. He found evidence of diseconomies of scale for two-year and four-year public, and four-year private college and university libraries, and economies of scale for two-year private college libraries. But the *R* square of the model for two-year private college libraries was low (0.50). He warned that careful interpretation of this finding was necessary.

Another noted author in the cost studies on academic libraries is Paul Kantor. Kantor (1981a, 1981b) developed cost functions for library operations in scientific and technical libraries. Kantor was also aware of the multiproduct nature of libraries. He thoughtfully included in-house material use, circulation, and reference queries as output variables. He concluded that the best-fit model clearly demonstrated economies of scale.

The previous studies utilized various econometric models. Some used the Cobb-Douglas production function. Some used the translog cost function. Others used general cost functions. As Cooper argued that "Both models (production function and cost function) are useful in determining whether scale economies exist"(1979, p. 66), Hayes, Cooper, Kantor, and Hammond recognized the multioutput nature of libraries and were able to incorporate various output variables into their studies while other researchers tended to use circulation as the sole indicator of output. DeBoer and Hammond used the translog cost function to estimate input substitution elasticities in addition to scale economies. Goddard and Feldstein also examined the relationship between the marginal cost and average cost.

The previous studies tended to focus on public libraries.[2] There probably are a few reasons why the earlier research focused on public libraries. First the data on public libraries were readily available at the local, state, and national level. Second, the policy incentive for studying public libraries was stronger because consolidating smaller-sized libraries could lead to cost savings. Third, it was believed that production activities of public libraries could be measured by a single output indicator. It was convenient to use a production function model with a single output measure.

This Study

This study argues that libraries in general, academic research libraries in particular, are information provision organizations providing multiproducts and multiservices. Their outputs are not homogenous and cannot be measured simply by a single output indicator. Traditional econometric methods, such as the production function, that can only be used to measure a single output are certainly not a sufficient measure of the production of academic research libraries. The multiproduct and multiservice

nature of libraries was not fully recognized by some researchers and multioutput variables were not incorporated in some previous studies. Hayes, Cooper, Kantor, and Hammond are among the very few researchers who were able to incorporate some multioutput variables into their studies.

To be able to provide more accurate estimates for library costs and economies of scale, this study took into account the multiproduct and multiservice nature of academic research libraries. This study is different from the previous studies in a number of ways. First, this study dealt with the cost function and scale economies of academic research libraries that belong to the Association of Research Libraries. As stated in the introduction of this article, academic research libraries have different missions and goals from general academic libraries and public libraries. One of the important goals of academic research libraries is to support research in universities. This research orientation demands that academic research libraries have extensive scholarly journal subscriptions, which consume a substantial part of library expenditures. This study treated serials as an independent output variable. Second, this study treated general library collections as outputs. A wide range of collection output variables were incorporated in the cost function. Third, this study added a new library service output variable, group presentations by librarians, to the cost function. Such data were not available to the previous research. Fourth, in this study, circulation was treated as only one of the service outputs.

Data, Measurements, and Model

All the data used in this study were collected from the ARL 1999–2000 survey of 112 academic research libraries (Association of Research Libraries, 2001). The model in this study incorporated eleven independent variables and one dependent variable. Because not all libraries had the data related to all of these variables, eighty- nine academic research libraries were included in the regression analysis.

A Multioutput Measure versus a Single Output Measure. In the previous studies on public libraries, most of the researchers used circulation as the sole output measure. Circulation was used as a single measure of output for public libraries because it was believed that circulation could capture most of the usage activities of public libraries and that the Cobb-Douglas production function was convenient to measure library output. Other reasons were that data on some of the output variables at that time were unavailable. The data collected unsystematically were considered as unreliable and inconsistent. In this study, the ARL data were used. The ARL has been systematically gathering data from its members for many years and its dataset has been widely used by researchers, library administrators, and practicing librarians. The ARL data are believed to be reliable and comprehensive although more detailed and more consistent data on academic research libraries need to be collected in the future.

This study treated the library collections and various library services as outputs. It included various types of library collection materials and library services. Library collections include volumes, serials, maps, graphs, videos, and audios. Library services are also an important part of library outputs. They include reference service, library seminars and workshops, interlibrary loans (borrowed and lent), and circulation or information delivery service. Library collections were used as output measures for a number of reasons. Library collections in this study were considered as final products of libraries. Although libraries do not directly create contents of library collections, such as contents of books and journal articles, and do not physically print these library materials, libraries do process them. The value added to materials acquired and purchased by libraries lies in the fact that these materials can be readily accessed, retrieved, and used by library patrons.

Volumes Held versus Volumes Added as an Output Measure. In this study, volumes held was used as an output measure instead of volumes added (which was used in two previous studies (Cooper, 1979, 1983)) because volumes added only measure the costs of volumes added to collections for one time period, typically one year. The volumes-added approach may be appropriate in other studies, but for this study, it was assumed that library users do not use just newly added volumes, they also use volumes purchased in the past. Maintaining existing volumes or entire bookstock is an ongoing process and involves a greater amount of staff time and effort and incurs more costs than newly added volumes and circulated items. The library collection management process generally includes assessing collections in terms of the needs of their patrons or communities, identifying, selecting, acquiring, classifying, cataloging, shelving, or storing all kinds of materials acquired and purchased by libraries. Many libraries assess their library collections in terms of their age and subject strengths and weaknesses so that librarians can make adjustment to support teaching and research or to compare with other libraries for collaboration purposes. Such an assessment requires searching and sorting entire collections and may take years to complete, but it is a necessary procedure to maintain relevant and useful collections. To put collections into a library, library staff must go through this collection management process. When library materials are in place for use, library staff also need to frequently evaluate collections, weed out those that have low values to make room for new purchased items, bind monthly and quarterly serial issues into annual volumes, repair damaged materials, replace missing items, and reshelve returned items. Other visible operating costs include electricity for lights and air conditioning. These costs are for entire collections not just for volumes added in a year. The use of the volumes-added approach in this study might tend to have biased estimates on the costs necessary to maintain entire collections.

In their recent study, Stephen R. Lawrence, Lynn Silipigni Connaway,

and Keith H. Brigham provided a convincing case that maintaining library collections is far more costly than one-time purchase costs using the ARL data (2001). For example, they demonstrated that unit cost for monographs is $47.78, yet their life-cycle costs are $343.03; unit cost for serials is $590.97, yet their life-cycle costs are $801.78; unit cost for videos is $15.70, but their life-cycle costs are $107.50; and unit cost for many government documents is 0 (free), but their life-cycle costs are $55.40 (p. 553). Life-cycle costs take into account ongoing expenses which include operating expenses, wages and salaries of staff, building and facilities, and facility maintenance. All of these costs are for one purpose: to make library collections readily available for use. Because the volumes-added approach might produce biased estimates on costs of maintaining existing volumes, the volumes-held approach was used in this study.

Volumes Held versus Circulation as an Output Measure. The use of circulation as a single output measure only considers costs of those items that are checked out from libraries. But whether or not library items are circulated is not very relevant from the fixed-cost perspective since most of the costs of making them available have already been incurred even before items are checked out. The function of circulation service is simply putting readily available items or finished products in the hands of users. The cost of circulation of a library item is only a small part of the total cost of processing this item and involves no more than staff time and equipment in the check-out process. Circulation was considered in this study as one of the service output measures.

Another reason for using volumes held as one of the output measures is that the use of circulation as an output measure for academic research libraries tends to ignore the fact that some parts of collections are not circulated and that users may use library materials in-house. Some materials, such as reference materials, must be used in the library. Many libraries do not have financial, human resources, and a mechanism to consistently track the in-house use of reference materials. Many academic research libraries also provide graduate students and faculty members with carrels where they can put the books for their learning, teaching, and research, and there is no need to check them out. Many users also use general library collections inside the library. They simply do not check them out. Circulation records do not reflect the usage of these library resources. The use of volumes held as an output variable covers all in-house use of library volume materials.

Serials as Output. Academic research libraries have extensive scholarly journal collections that are important library assets for learning, teaching, and research. Journals make up a large portion of serials. In general, journals in virtually every academic research library, like reference materials, are not circulated items and are not recorded for use. Some libraries may record current journal usage, but they are unlikely to record usage of back

issues on a regular and continuous basis since it would incur too much cost related to staff time and efforts. Some academic libraries may track the use of current journals by requiring patrons not to put them back on the shelves so that library staff can record which journals have been used and how often. To track the use of back issues of tens of thousands of serials is extremely difficult, if not impossible. Back issues of journals are used a great deal by both faculty and students because they contain accumulated research findings in the past and are indispensable for research and teaching. To study scale economies of academic research libraries without including serials would miss a critical part of academic research library output. Serials collections consume a large portion of academic research libraries' budgets and were treated as an important output variable in this study.

Other Library Collections as Output. In addition to volumes held and serials, other library collections, such as audios, videos, and maps were also treated as library outputs since they are different kinds of materials from volumes and serials. The prices of these materials differ from those of volumes and serials and do incur costs related to staff time and library equipment in order to make them readily available for use. In general, some library materials such as large-sized maps and some audio and video items may well be in-house use materials. Circulation records do not record such use of library collections. Although it is true that only a small portion of library collections is used at a certain time period, it does not necessarily mean that only that portion incurs costs.

Library Services as Output. Libraries provide various types of services including reference service, instruction in the form of library workshops and seminars, borrowing items through interlibrary loan for local users and lending items through interlibrary loan to external users, and circulation service. These services incur costs in terms of staff time, facilities, and equipment necessary to carry out these activities. Reference service can be measured by reference transactions, and library workshops and seminars are measured by library group presentations. Interlibrary loan and circulation data are also readily available from the ARL dataset.

Library Costs. Total library expenditures were used as a measure of total library costs on library materials, staff, binding, and other operating activities. Library materials expenditures consist of costs for monographs, serials, other materials such as maps, audio and visual items, and the items other than materials such as bibliographic utilities. Total salary expenditures include those for professional staff, non-professional staff, and student assistants.

The Model. The general form of total cost function is the following:

$TC = f(V, S, U, D, M, G, P, R, B, L, C)$;

TC, the total cost, is a function of a wide range of library outputs. The

letters *V, S, U, D, M, G, P, R, B, L,* and *C* represent library collections and service outputs and can be written as:

1. $TC = A\ V^{\beta 1}\ S^{\beta 2}\ U^{\beta 3}\ D^{\beta 4}\ M^{\beta 5}\ G^{\beta 6}\ P^{\beta 7}\ R^{\beta 8}\ B^{\beta 9}\ L^{\beta 10}\ C^{\beta 11}$

Taking the natural log of both sides to derive the following cost equation:

2. $\ln TC_i = \ln A + \beta_1 \ln V_i + \beta_2 \ln S_i + \beta_3 \ln U_i + \beta_4 \ln D_i + \beta_5 \ln M_i + \beta_6 \ln G_i + \beta_7 \ln P_i + \beta_8 \ln R_i + \beta_9 \ln B_i + \beta_{10} \ln L_i + \beta_{11} \ln C_i + \varepsilon_i$

Where

i indexes individual institutions ($i = 1, \ldots, N$),
N is the total number of observations,
TC is the total cost,
A is the constant,
$\beta_1, \beta_2, \beta_3, \ldots, \beta_{11}$ are the coefficients,
ε is the statistical noise or the error term,
It is specified that:
$A > 0, \beta_1 > 0, \beta_2 > 0, \ldots, \beta_{11} > 0.$

Library collections outputs:

V is the total number of volumes held,
S is the total number of serials,
U is the total number of audio items,
D is the total number of video items,
M is the total number of maps,
G is the total number of graphs,

Library service outputs:

P is the total number of group presentations,
R is the total number of reference transactions,
B is the total number of interlibrary loans borrowed,
L is the total number of interlibrary loans lent,
C is the total number of circulated items.

Findings and Analysis

Table 1 shows that the adjusted R square of the model is 0.8, indicating that about 80 percent of the variance of the dependent variable can be explained by the model. The adjusted R square is high enough not to reject the model. The t statistics show that the coefficients of volumes held and serials, very important parts of library collections, are statistically significant at a very high level (.0001 and .0003 respectively). The coefficient of group presentations, part of library services, is also statistically significant. The coefficients of other variables are not statistically significant except for the coefficient of maps with a negative sign, which does not make any sense, and needs to be disregarded. As Hammond pointed out, "it is not practical to include all the identifiable dimensions of output. In addition, some

Table 1. Regression Results from the Natural Log Linear Model, Dependent Variable is *ln TC.*

Variable	Parameter Estimate	Standard Error	*t* Value	Sig. Level
INTERCEPT	5.134733	0.96568951	5.317	0.0001
ln V	0.515291	0.10586920	4.867	0.0001
ln S	0.285986	0.07618958	3.754	0.0003
ln U	0.019001	0.02414557	0.787	0.4337
ln D	0.030656	0.02543264	1.205	0.2318
ln M	–0.031324	0.01599503	–1.958	0.0538
ln G	–0.003305	0.01098661	–0.301	0.7643
ln P	0.093980	0.04482646	2.097	0.0393
ln R	0.024787	0.04541695	0.546	0.5868
ln B	0.030396	0.04270257	0.712	0.4787
ln L	–0.042342	0.05273819	–0.803	0.4245
ln C	0.005471	0.05031736	0.109	0.9137
Adjusted $R^2 = 0.8$				
Cases = 89				

dimensions may not be easily quantified (1999, p. 274)." With three variables that measure important parts of libraries' collections and service outputs, the regression results should be considered satisfactory.

To measure scale economies, the total coefficients of the independent variables are summed. If the function coefficient is larger than 1, then diseconomies of scale exist. If the function coefficient is smaller than 1, then economies of scale exist. If the function coefficient equals 1, then constant returns to scale exist. The function coefficient ($\beta_1 + \beta_2 + \beta_3, \ldots, + \beta_{11}$) is found to be 0.928597 or 0.93, indicating that small economies of scale exist in academic research libraries.

Comparing With Previous Research

The studies on scale economies of libraries have reached different conclusions: diseconomies of scale, economies of scale, and constant returns to scale. The mixed findings should not be surprising for a number of reasons. First, studies were conducted on a wide range and diverse groups of libraries, including public libraries, scientific and technical libraries, two-year and four-year academic libraries, private and public college libraries, and academic research libraries. Second, the data were gathered at different levels. Some studies focused on libraries within one state. Some studies gathered data from a region or a number of states. Others used the national data. Third, econometric models used in the studies vary from study to study. Some used the production function. Some used the translog cost function. Others used general cost functions. Fourth, the variables used in the models vary from study to study. Some used a single output variable. A few used multiple output variables. The production coefficient which measures scale

economies is very sensitive to the number of variables used and which variables are used. Table 2 shows such a diversity of studies on scale economies of libraries in terms of model, output and input variables, and findings.

Table 2. Comparison Between Studies on Scale Economies of Libraries.

Investigator	Variable	Coefficients	Model
Black	Inputs		
1969	1. Bookstock	0.167	Log Production
p. 595	2. Labor	0.833	Function
	Sum of the Coefficients	1.0	
	Conclusion	Constant Returns to Scale	
Goddard	Inputs		Log Production
1973	1. Bookstock	0.486	Function
p. 198	2. Labor	0.160	
	3. Materials	0.111	
	4. Capital	0.337	
	Sum of the Coefficients	1.076[3]	
	Conclusion	Slight Increasing Returns to Scale	
Feldstein	Outputs		Total Cost
1976	1. Circulation only.		Function
	Observed Marginal Cost and Average Cost		Average Cost Function
	Conclusion	The Library System had Diseconomies of Scale; but Some Individual Libraries had Economies of Scale.	
Cooper	Public Library Outputs		Log Cost
1979	1. ILL Borrowed	0.551	Function
p. 74	2. ILL Lent	–0.00058	
	3. Reference Transactions	–0.0062	
	4. Circulation	0.017	
	5. Volume Added	0.467	
	Sum	1.028 (FY1975/76)	
	Conclusion	Constant Return to Scale	
1983	Two-Year Public College Libraries		ln Cost
p. 216	Outputs		Function
	1.Volumes Added	0.4218	
	2. Reference Transactions	0.0921	
	3. Circulation	0.2705	
	4. Hours Opened	0.5335	
	5. Interlibrary Loan Lending	0.0124	

Table 2. Con't.

Investigator	Variable	Coefficients	Model
	6. Interlibrary Loan Borrowing	0.0082	
	Sum	1.3	
	Conclusion	Diseconomies of Scale	
	Two-Year Private College Libraries		
	Outputs		
	1.Volumes Added	0.3301	
	2. Reference Transactions	0.0877	
	3. Circulation	0.1271	
	4. Hours Opened	0.1817	
	5. Interlibrary Loan Lending	0.0691	
	6. Interlibrary Loan Borrowing	0.0061	
	Sum	0.80	
	Conclusion	Economies of Scale	
	Four-Year Public College and University Libraries		
	Outputs		
	1.Volumes Added	0.5554	
	2. Reference Transactions	0.0957	
	3. Circulation	0.1374	
	4. Hours Opened	0.3426	
	5. Interlibrary Loan Lending	0.0635	
	6. Interlibrary Loan Borrowing	0.0410	
	Sum	1.2	
	Conclusion	Diseconomies of Scale	
	Four-Year Private Collection and University Libraries		
	Outputs		
	1.Volumes Added	0.5195	
	2. Reference Transactions	0.1171	
	3. Circulation	0.2121	
	4. Hours Opened	0.1706	
	5. Interlibrary Loan Lending	0.0708	
	6. Interlibrary Loan Borrowing	0.0359	
	Sum	1.1	
	Conclusion	Diseconomies of Scale	
Kantor 1981	Outputs		ln Cost Function
	1. In-House Materials Use	0.11	

Table 2. Con't.

Investigator	Variable	Coefficients	Model
Part II	2. Circulation	0.32	
p. 149	3. Reference Queries Received	0.32	
	Sum	0.75	
	Conclusion	Economies of Scale for Scientific and Technical Libraries	
DeBoer	Inputs		Translog Cost
1992	1. Bookstock		Function
p. 266	2. Supplies/Service		
	3. Books		
	Circulation Level		
	3,633	0.856	
	14,209	0.898	
	55,409	0.940	
	216,075	0.982	
	842,610	1.024	
	Conclusion	Economies of Scale for Smaller-Sized Public Libraries but Constant Returns to Scale for Larger-Sized Public Libraries.	
Hammond	Outputs		Translog Cost
1999	1. Books	0.4682	Function
p. 287	2. Audio/Visual Materials	0.0773	
	3. Inquiries	0.1303	
	Conclusion	Increasing Returns to Scale. Diseconomies of Scope for the Average British Public Library.	
Liu	Outputs		In Cost
2002	Library Collections Outputs		Function
	1. Volumes Held	0.515291	
	2. Serials	0.285986	
	3. Audio	0.019001	
	4. Video	0.030656	
	5. Maps	–0.031324	
	6. Graphs	–0.003305	
	Library Service Outputs		
	1. Group Presentations	0.093980	
	2. Reference Transactions	0.024787	

Table 2. Con't.

Investigator	Variable	Coefficients	Model
	3. Interlibrary Loans Borrowed	0.030396	
	4. Interlibrary Loans Lent	–0.042342	
	5. Circulations	0.005471	
	Sum	0.928597	
	Conclusion	Slight Economies of Scale for Academic Research Libraries.	

The findings of this study show that volumes held and total number of serials are significant output variables. The coefficients of these two variables are statistically significant at a very high level. A correlation analysis (not presented here) indicated that each of them is highly correlated with the dependent variable. The correlation between volumes held and the total cost was found to be 0.89 and the correlation between serials and the total cost is 0.82, confirming that they are good indicators of outputs of academic research libraries. Group presentations that reflect the total number of library workshops and seminars offered is also a significant output. The data related to group presentations were collected only in recent years by the ARL.

This study was not concerned with input substitution elasticities under the assumption that volumes held, serials, and group presentations, three major output variables, are not likely to be substitutes. Academic research libraries do not buy more books using serials expenditures simply because books are cheaper. Journals provide up-to-date research findings. They are critical for research and teaching and cannot be replaced by books. It is also unlikely that libraries would reduce the number of library instructors teaching library workshops and use the savings to purchase journals due to higher labor costs. Library instruction is indispensable training and education for students. It provides students with the information competency they need to effectively access, retrieve, evaluate, and use information.

The regression results of this study also show that circulation is not a good measure of library output since it is not statistically significant. The correlation analysis also revealed that the correlation between circulation and the total cost is only 0.50, lower than volumes held (0.89), serials (0.82), group presentations (0.54), and audios (0.57). Given the nature of academic research libraries, a great deal of library materials, such as serials, reference collections, and materials in carrels, is used in-house. Circulation records do not reflect such usage.

The results in a separate regression that used volumes added as an output variable in place of volumes held with other variables unchanged showed that the adjusted R square decreased to 0.76 from 0.8 and the sum of the coefficients decreased to 0.86 from 0.93. As expected, using volumes

added for the purpose of this study could have overestimated economies of scale because volumes added do not take into account staff and operating costs involved in maintaining existing collections.

Summary and Conclusion

This study examined scale economies of academic research libraries and reviewed the research literature on economies of scale in various libraries. It argued that academic research libraries are information provision organizations providing multiproducts and multiservices. The total cost function was developed and the natural log linear model was proved to be the best fit of the data. A wide range of collections and service outputs were incorporated into the cost function to reflect this information provision function. Library outputs were measured by library collections, including volumes held, serials, audio and video materials, maps, graphs; and library services, including library workshops and seminars in the form of group presentations, reference transactions, circulation service, and interlibrary loan services. Three major output variables—volumes held, serials, and group presentations—stand out in terms of statistical significance. It was found that slight economies of scale exist in academic research libraries.

Previous research has made an important contribution to our understanding of scale economies of libraries and developed various methodologies that can be used in the later studies. But previous research also has some limitations. One of the obvious limitations is applying the Cobb-Douglas production function, which is normally used to measure a single output, to libraries that provide a wide range of outputs and services. This problem could be overcome by applying the production function to cross-section data at the departmental or division level. But this did not happen probably because of the lack of understanding of library operations by some researchers and the lack of consistent and reliable data related to library outputs, which frustrated researchers. The ARL has been collecting data for academic research libraries for many years. The statistics gathered have increasingly reflected the multiproduction nature of academic research libraries. For example, data on group presentations measured an important service activity of academic research libraries. More work needs to be done in the future to collect data on capital, labor, and costs at the department or division level so that cross-section department or division production functions can be used to measure the multiproduct activities of academic research libraries.

It is hoped that this investigation on scale economies in academic research libraries will provide some new insights into the existing literature in terms of understanding of scale economies for libraries and output variables used in the total cost function. Because scale economies are very sensitive to the number and the nature of output variables used in the regression model, it is very important for researchers to carefully select output

variables. Good output indicators should reflect the true costs of libraries' outputs.

ACKNOWLEDGMENTS

I would like to thank Dr. Richard E. Quandt, an emeritus professor of economics at Princeton University, and Dr. Robert M. Hayes, an emeritus professor of library and information science at UCLA, for their comments and suggestions on an earlier draft of this article. I am alone responsible for all possible errors. This article was first published in the September 2002 issue of *College & Research Libraries.*

NOTES

1. American Library Association. ALA Library Fact Sheet Number 1 contains various library survey results provided by ALA (Chicago: American Library Association, 2001). Retrieved March 29, 2002, from http://www.ala.org/library/fact1.html.
2. See works by Black; Goddard; Feldstein; Hayes; Cooper; DeBoer; and Hammond.
3. The function coefficients do not add up to the sum although it is close.

REFERENCES

American Library Association. (2001). ALA Library Fact Sheet Number 1. Chicago: American Library Association. Retrieved March 29, 2002, from http://www.ala.org/library/fact1.html.

Association of Research Libraries. (2001). *ARL Statistics* 2000–2001. Washington, DC: Association of Research Libraries. Retrieved December 10, 2001, from http://www.arl.org/stats/arlstat/index.html.

Bauer, P. W. et al. (2000). Optimal use of scale economies in the Federal Reserve's currency infrastructure. *Federal Reserve Bank of Cleveland Economic Review, 36*(3), 13–27.

Baumol, W. J., & Marcus, M. (1973). *Economics of Academic Libraries.* Washington, DC: American Council on Education.

Baumol, W. J. et al. (1969). The costs of library and informational services. In Knight, D.M. & Nourse, E. S. (Eds.), *Libraries at large* (pp. 168–227). New York: R.R. Bowker.

Black, S. W. (1969). Library economics. In Knight, D. M. & Nourse, E. S. (Eds.), *Libraries at large.* New York: R.R. Bowker, 590–598.

Chakraborty, K., Biswas, B., & Lewis, C. W. (2000). Economies of scale in public education: An econometric analysis. *Contemporary Economic Policy, 8*(2), 238–248.

Cohn, E., Rhine, S. L.W., & Santos, M. C. (1989). Institutions of higher education as multiproduct firms: Economies of scale and scope. *Review of Economics and Statistics, 71*(2), 284–290.

Cooper, M. D. (1979). The economics of library size: A preliminary inquiry. *Library Trends, 28*(1), 63–78.

Cooper, M. D. (1983). Economies of scale in academic libraries. *Library and Information Science Research, 5,* 207–219.

DeBoer, L. (1992). Economies of scale and input substitution in public libraries. *Journal of Urban Economics, 32*(2), 257–268.

Feldstein, K. F. (1976). "The Economics of Public Libraries." Ph.D. diss., Massachusetts Institute of Technology, Boston.

Goddard, H. C. (1973). Analysis of social production functions: The public libraries. *Public Finance Quarterly, 1*(2), 191–204.

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*(3), 271–295.

Hayes, R. M. (1979). The management of library resources: The balance between capital and staff in providing services. *Library Research, 1,* 119–142.

Hayes, R. M. (1981). The distribution of use of library materials: Analysis of data from the University of Pittsburgh. *Library Research, 3*(3), 215–260.

Hayes, R. M. & Borko, H. (1983). Mathematical models of information system use. *Information Processing and Management, 19*(3), 173–186.

Kantor, P. (1981a). Levels of output related to cost of operation of scientific and technical libraries: Part I. Techniques and cumulative statistics. *Library Research, 3*, 1–28.

Kantor, P. (1981b). Levels of output related to cost of operation of scientific and technical libraries: Part II. A capacity model of the average cost formula. *Library Research, 3*, 141–154.

Koshal, R. K. & Koshal, M. (2000). Do liberal arts colleges exhibit economies of scale and scope? *Education Economics, 8*(3), 209–211.

Lawrence, S. R., Connaway, L. S., & Brigham, K. H. (2001). Life cycle costs of library collections: Creation of effective performance and cost metrics for library resources. *College & Research Libraries, 62*(6), 541–544.

Liu, L. G. (2001). The contribution of library collections to prestige of academic programs of universities: A quantitative analysis. *Library Collections, Acquisitions, and Technical Services, 25*(1), 49–65.

Mansfield, E. (1997). *Microeconomics: Theory and Applications.* 9th ed. New York: W. W. Norton, 191.