Economies of Scale in Academic Libraries

MICHAEL D. COOPER*

University of California, Berkeley

This paper examines the relationship between costs and output in more than 3,000 United States college and university libraries. Major differences in workload and expenditures are found between two-year and four-year public and private institutions. Analyses show that average costs rise more rapidly than output for all types of libraries except two-year private college libraries.

INTRODUCTION

Academic libraries are heterogeneous in their characteristics. This paper explores those differences by type (public, private, two-year, four-year) and by size of the libraries. The analysis briefly reviews comparative studies of the cost characteristics of academic libraries, presents a summary of the descriptive statistics of college and university libraries included in this study, and analyzes the correlation and partial correlation between some of the descriptive statistics. Emphasis throughout is given to cost relations and the variations in costs by type of library. The paper concludes by showing that except for two-year private college libraries, the average cost of providing a unit of service rises faster than output thus eliminating any possibility of economies of scale in library operations. Economies of scale occur when the average costs per unit of output rises less rapidly than output, and diseconomies occur when average cost rises faster than output. The middle case, where average cost rises at the same rate as output is termed constant returns to scale.

*The author acknowledges Sandy Warmington for her assistance at the initial stages of this project, to Betsy Flores, for her research assistance, and to Kitty Whiteside for her data processing assistance. An anonymous referee made a number of valuable suggestions to this paper. Correspondence and requests for reprints should be addressed to Michael D. Cooper, School of Library and Information Studies, University of California, Berkeley, CA 94720.

Previous Studies

In an earlier paper Cooper (1979) reviewed the problems of measuring the size of libraries and the literature on economies of scale in private and public organizations. An empirical investigation was made to determine whether economies of scale were present in public libraries in California. The methodology established a statistical relationship between the output produced by libraries and the cost of producing that output. Based on an examination of the cost-output relationship, it was determined that for public libraries in California the cost increased as output increased, and that there were almost constant returns to scale.

Several subsequent papers have addressed the question of cost functions and economies of scale. Hayes (1979) analyzed public libraries in California, Illinois, Ohio, Missouri, and Wisconsin. His approach was to determine whether a Cobb-Douglas form of a production function could characterize the cost relations in libraries. In this model a division is made between labor and capital used in a library. Hayes used the size of library collections as a surrogate for capital costs and reader service staff as a surrogate for labor costs. He then performed a regression analysis on the data from selected public libraries in the five states and found a relatively consistent mix among states in the proportion of labor and capital used.

Kantor (1980) collected primary data on the cost of library operations and used that data to develop, among other things, a cost-output model. The purpose of the model was to ascertain whether economies of scale existed. His model related total operating expenses to three measures of output: patron hours of in-library use of materials, circulation, and reference queries received. Using a sample of 65 libraries he found that the best fit to the data was obtained from a logarithmic form of an equation and that there were economies of scale in his sample.

Lawrence (1981) developed cost functions for the nine campus libraries of the University of California system. He used a model in which total library expenditures were a function of circulation, interlibrary lending, interlibrary borrowing, number of items withdrawn from the collection, volumes added to the collection, number of public service points, total hours the public service points were open during a year, and the year of the observation. He found economies of scale in processing additional materials, in the circulation function, and in weeding materials from the collection.

DATA SOURCE

The data used in this paper's analysis of cost trends among academic libraries comes from statistics gathered by the National Center for Education

Statistics (NCES) for a survey of college and university libraries conducted for the year 1977. The institutional data is reported by Beasley (1980), and summary data is reported in Beasley (1981). Although intended for a different purpose, Carpenter's report (1981) provides additional summary statistics from the survey.

COMPARISON OF DESCRIPTIVE STATISTICS FOR ACADEMIC LIBRARIES

There are major differences between the four categories of libraries selected for this analysis. The categories are college and university libraries found within two-year public, and two-year private, four-year public, and four-year private institutions. Out of 3057 institutions included in this analysis, 908 were two-year public, 241 were two-year private, 553 were four-year public, and 1355 were four-year private.¹

Table I presents the mean values for some of the variables that can be used to describe the characteristics, size and output or workload of the libraries. The standard deviations of the mean values are also given and tend to be rather large, indicating high variability in the data.

The table shows that the average number of volumes held by all academic libraries was 184,083 in 1977 but it varied from a low of 22,565 for two-year private institutions to 526,042 for four-year public institutions. Wide variations can also be observed in many other variables. For example, the average number of items circulated for all libraries was 85,848, while four-year public institutions had 193,468 items circulated, and two-year public had 29,319 items circulated. The table also shows that the major burden in interlibrary lending is being carried by four-year public libraries, but they also do the most interlibrary borrowing. In general, the four-year public institutions had the largest, and the two-year private institutions had the smallest, mean values for every category.

Table II presents data on staff size and costs for the academic libraries. The mean Full Time Equivalent (FTE) number of staff in all types of libraries varied from 10 to 41, and mean salary expenditures varied from \$21,520 to \$575,000. Overall operating expenditures averaged \$412,677 for all libraries, but for two-year public institutions they were \$197,790; for two-year private institutions, \$41,528; four-year public, \$1,226,079 (by far the largest); and four-year private, close to \$290,000. In one respect the figures are remarkably similar. Dividing the mean salary expenditures by mean

'The number of libraries falling into each category in this analysis does not correspond exactly to the numbers reported by the National Center for Education Statistics, primarily because institutions were eliminated in this analysis when all reported data were either blank or zero.

TABLE I
Characteristics, Size and Workload of Academic Libraries, 1977

		All Libraries	2-Year Public	2-Year Private	4-Year Public	4-Year Private
Volumes Held	Mean	184083	40770	22565	526042	169287
	St. Dev.	476209	34009	22991	738484	471423
	N	3057	908	241	553	1355
Volumes Added	Mean	7518	3022	1285	20782	6208
	St. Dev.	15562	2799	2107	24860	13933
	N	3050	908	239	553	1350
Interlibrary	Mean	1637	149	115	4898	1098
Lending	St. Dev.	8903	474	394	15775	6733
•	N	2435	690	102	544	1099
Interlibrary	Mean	849	282	118	2307	673
Borrowing	St. Dev.	1937	626	404	3103	1608
· ·	N	2724	855	147	548	1174
Reference	Mean	764	589	271	1713	563
	St. Dev.	1876	1196	850	3001	1622
	N	2677	818	211	502	1146
Circulation	Mean	85848	29319	10705	193468	47300
	St. Dev.	163272	38208	33361	289979	126188
	N	3015	903	229	551	1332
Hours Open	Mean	74	65	59	86	78
per week	St. Dev.	19	12	17	15	21
-	N	3054	907	241	553	1353

Source: Derived by the author from the machine-readable version of the data published by Beasley (1980).

TABLE II
Staff and Expenditures for Academic Libraries, 1977

		All Libraries	2-Year Public	2-Year Private	4-Year Public	4-Year Private
Total FTE	Mean	22	15	10	41	21
Staff	St. Dev.	52	19	15	71	58
0.00.1	N	1994	547	151	362	934
Total Salary	Mean	202032	110860	21520	575000	138964
Expenditures	St. Dev.	473782	123005	26878	755984	443876
	N	2985	906	226	551	1302
Total Material	Mean	130696	45986	10662	408532	94884
Expenditures	St. Dev.	291667	56678	10964	472143	241654
•	N	3028	908	241	552	1327
Total Other	Mean	33507	17655	4109	87646	25818
Expenditures	St. Dev.	91392	30396	12291	125417	100422
	N	2881	881	209	548	1243
Total Operating	Mean	412677	197790	41528	1226079	289999
Expenditures	St. Dev.	950135	214752	54719	1531759	841277
	N	3049	908	241	553	1347

Source. See Table I.

operating expenditures it is found that the proportion of money spent on salaries is quite similar between library types. It varies from 47% for four-year public college and university libraries, to 48% for four-year private, to 52% for two-year private, to 56% for two-year public college libraries. Again, the same pattern holds in this table as in the previous one: four-year public institutions have the largest mean values for expenditures, and two-year private, the smallest.

CORRELATION ANALYSIS

Comparisons of the characteristics of college and university libraries can also be performed through correlation of some of the measures of library output. The correlations are interesting in their own right because they attempt to explain the affect that one library output variable has on another. For example, the presence or absence of correlation between the size of the collection and the degree of circulation could show how the size of the collection affects circulation. The purpose of discussing correlation analysis here is as a precursor to the regression analysis of cost versus output. In that analysis it will be important to know the extent to which certain variables uniquely characterize library output and the extent to which they measure the same aspect of library output as another variable.

Tables III through VI present the simple Pearson correlations between eight selected variables. A number of points arise in comparing these data. Relative to the other three types of libraries, the correlations between variables for two-year private college libraries are the weakest, perhaps reflecting unstable patterns of funding. With four-year public institutions, there are relatively strong correlations between the number of reference transactions and volumes held, and volumes added and circulation.

TABLE III
Correlations for Two-Year Public College Libraries, 1977

	Volumes Added	Inter- Library Lending	Inter- Library Borrow.	Ref- erence	Hours Open	Circu- lation	Total Operating Expense
Volumes Held	.67	.18	.11	.32	.33	.67	.68
Vols. Added		.09	.07	.22	.38	.50	.66
ILL			.42	.08	.02	.12	.10
ILB				.02	.02	.07	.06
Reference					.08	.30	.26
Hours Open						.37	.39
Circulation							.70

TABLE IV
Correlations for Two-Year Private College Libraries, 1977

	Volumes Added	Inter- Library Lending	Inter- Library Borrow.	Ref- erence	Hours Open	Circu- lation	Total Operating Expense
Volumes Held	.48	.27	.13	.06	.26	.68	.85
Vols. Added		.20	.15	.22	.26	.40	.34
ILL			.79	.00	.10	.24	.36
ILB				03	.05	.10	.17
Reference					.10	01	.07
Hours Open			4			.13	.26
Circulation							.69

TABLE V
Correlations for Four-Year Public College and University Libraries, 1977

	Volumes Added	Inter- Library Lending	Inter- Library Borrow.	Ref- erence	Hours Open	Circu- lation	Total Operating Expense
Volumes Held	.88	.52	.49	.55	.39	.88	.93
Vols. Added		.43	.48	.57	.45	.86	.91
ILL			.58	.27	.22	.47	.52
ILB				.25	.37	.46	.54
Reference					.18	.58	.62
Hours Open						.40	.42
Circulation							.89

TABLE VI
Correlations for Four-Year Private College and University Libraries, 1977

	Volumes Added	Inter- Library Lending	Inter- Library Borrow.	Ref- erence	Hours Open	Circu- lation	Total Operating Expense
Volumes Held	.93	.59	.55	.38	.21	.84	.97
Vols. Added		.48	.52	.39	.24	.81	.95
ILL			.48	.18	.08	.51	.58
ILB				.19	.20	.51	.56
Reference					.12	.37	.40
Hours Open						.19	.22
Circulation							.86

The strongest most consistent correlation across all four types of libraries occurs between the number of items circulated and the number of volumes held and added. However, further analysis of the partial correlations of the variables reveals a different underlying pattern.

The partial correlation between two variables is a way of measuring the strength of the relation between them while controlling for a third. In the present analysis partial correlation was performed between circulation and volumes added, controlling for volumes held. The results show that uniformly across all types of libraries, the meaningful correlation is not circulation with volumes added but with volumes held. When the effect of volumes held is controlled, the correlation between circulation and volumes added drops to a very small quantity.²

An important concept in the analysis that follows is a measure of the output of a library. Clearly, this is difficult to define, let alone quantify.³ As an approximation to the output produced, the approach taken is to use certain measurable library variables that are found in the NCES data base. These include volumes added to the collection, interlibrary loans, interlibrary borrowing requests, reference transactions, number of hours the library is open per year, and number of circulation transactions.

The inclusion of volumes added as an output measure is an attempt to include a variable which will represent the amount of technical processing effort involved in acquiring and processing items for the library.⁴

Table VII explores whether there are strong interdependencies between the library output variables. One is interested in insuring that each variable uniquely measures a particular aspect of library output. Through the use of partial correlation analysis the extent of interdependencies between the six output measures can be examined.

A comparison of Table VII with Tables III through VI shows some of the interrelationships. In general, there is little change in the correlation between volumes added to a collection and circulation when the number of reference transactions is controlled.⁵

There is also a correlation pattern in four-year versus two-year institutions. Relatively little drop-off occurs in the partial correlations of volumes added and reference controlling for circulation, and reference and circula-

²The simple correlation between circulation and volumes added for two-year public, two-year private, four-year public and four-year private libraries is .50, .40, .86, and .81 respectively. Controlling for volumes held, the partial correlations for the same pairs are .10, .12, .40, and .14 respectively.

³See Cooper (1979) for a discussion of workload measurement.

'Kantor (1980, pp. 2-5), in a review of previous work in the area of economies of scale, questions whether volumes added is an output measure and whether it should be omitted from the computation. If one differentiates between work performed by library staff that *directly* affects the user, and all work performed by the staff, then counting circulation and reference transactions makes more sense with the former definition. Technical processing activities are a major step in providing service and consume a major proportion of library resources. Including volumes added is a means of reflecting the technical processing work performed. The other variables measure public service work performed.

⁵For example, the simple correlation between the two variables, volumes added and circulation is .50, .40, .86, and .81 respectively for the four library types. The partial correlations controlling for reference are .47, .42, .80, and .78. Very little change is evident between the pairs (e.g., .50 and .47; .40 and .42).

TABLE VII
Partial Correlation Analysis of Selected
College and University Library Statistics, 1977

Variable Pairs Correlated	Variable	Partial Correlation Coefficient				
	Controlled	2-Year Public	2-Year Private	4-Year Public	4-Year Private	
Volumes added and Reference	Circulation	.09	.25	.16	.15	
Volumes added and Circulation	Reference	.47	.42	.80	.78	
Reference and Circulation	Vols. added	.22	11	.22	.12	

Note. The number of observations used to derive the partial correlations for each of the four groups is 616, 80, 492 and 968 respectively.

tion controlling for volumes added, as compared to the simple uncontrolled correlations for two-year institutions. The opposite holds for four-year institutions.

The implications are that some interdependencies are present between the output variables but they are not uniform across types of libraries. Only for four-year institutions are they significant. There is some doubt whether multicollinearity exists between circulation and volumes added that should exclude one of them from measuring output for four-year institutions. There are two problems. First, each of the two variables measures something very different and thus each should be included on the basis of knowledge of the system. Second, if one or the other were to be excluded, it is not clear which it should be. For these reasons both are included as output measures.

REGRESSION ANALYSIS

The previous comparisons of descriptive statistics of the four categories of college and university libraries concentrated on differences in the magnitude of various measures of output and cost. Correlation analysis pointed up the degree of differences in the relation of one variable to another. This section analyzes the relationship between the costs or total operating expenses of libraries and the measures of output discussed in the previous section. The purpose of the analysis is to determine the statistical relationship between cost and output so that inferences can be made about whether there is an optimal size for libraries.

As described in a previous article (Cooper, 1979) there are three general patterns to the relation between average cost of producing a unit of output and the output of libraries. The first is a linear relation in which as the

output increases the average cost of performing that output increases. The second occurs where as the output increases, average costs first increase, reach a peak and then decline. This situation represents economies of scale: average costs decline after a point, reflecting greater efficiencies after a certain size is reached. The third model is one in which average costs decline and continue to decline as the output of a library increases. This model does not have much face validity in a library setting, given what we know about the cost structure of the institution. It does have applicability in industries like electrical power generation where there are large fixed costs of constructing plants and as output increases those fixed costs are amortized over more units of power output.

The equation used to verify the cost relationship that exists between types of libraries had the form:

$$\ln Y = \ln a + b_1 \ln X_1 + b_2 \ln X_2 + b_3 \ln X_3 + b_4 \ln X_4 + b_5 \ln X_5 + b_6 \ln X_6$$

where Y is the total operating expense and X_1 represents volumes added, X_2 represents reference transactions, X_3 -circulation, X_4 -hours opened, X_5 -interlibrary lending, and X_6 -interlibrary borrowing. The a's and b's in the equation are constants and 1n is the natural logarithm of the value. This equation has the property that depending on the value of the constants b's, it can represent a straight line or a curve concave or convex to the origin.

Ordinary Least Squares regression analysis was performed on the data from each of the four library groups, and the results are given in Table VIII. Between types of libraries the explantory power of the log-log equation varies, it is determined by the value of R^2 . For example, for two-year public libraries, it is .65; for two-year private, .50; four-year public, .88; and four-year private, .84.

The fit of the equation to the data for two-year private college libraries is the least strong of the four: because of the low R^2 and b's that are not significant, it should be treated with caution. The b_1 values in the equations for the four-year institutions are all significant at the $\alpha = .05$ level as measured by the t-statistic. The coefficients of the two-year public equation that are not significant are interlibrary lending and borrowing. These two variables do not have a large quantity of resources devoted to them, nor do they generate a large volume of activity in this type of institution.

Using the Durbin-Watson statistic, tests for autocorrelation were performed to determine whether there was any interdependence among successive values of the disturbance term in the regression equation. The equation showed no autocorrelation for any of the four categories of libraries.

'The low number of observations included in the two-year private group reflects the fact that all observations for all variables must be present for the institution to be included in this type of analysis. Poor reporting, and presumably recording as well, on the part of these libraries reduced the number from a possible 241 to 76.

TABLE VIII					
Regression Results for College and University Librar	ries				

	Two-Year Public	Two-Year Private	Four-Year Public	Four-Year Private
In a	1.0371	5.2143	2.4622	2.9761
	(1.08)	(3.08)	(2.44)	(6.66)
b_1	.4218	.3301	.5554	.5195
Volumes Added	(13.84)	(3.70)	(19.08)	(24.22)
b_1	.0921	.0877	.0957	.1171
Reference	(4.54)	(1.71)	(5.36)	(8.56)
b_3	.2705	.1271	.1374	.2121
Circulation	(13.29)	(3.25)	(6.60)	(12.60)
b_4	.5335	.1817	.3426	.1706
Hours Open	(4.15)	(.80)	(2.66)	(2.92)
b,	.0124	.0691	.0635	.0708
ILL	(.97)	(2.32)	(5.26)	(7.47)
b_{ϵ}	.0082	.0061	.0410	.0359
ILB	(.53)	(.20)	(2.35)	(3.35)
R ²	.65	.50	.88	.84
Durbin-Watson	1.92	2.26	1.93	1.74
Degrees of Freedom	612	76	488	963

Note. The value of the t-statistic is given in parentheses below the value of each coefficient. The F-statistic is significant at the $\alpha.05$ level for all equations.

ECONOMIES OF SCALE

The equation used to explain cost variations has the desirable property that by examining the sum of the b_1 values the existence of economies of scale can be established. If the sum of the b_1 's is less than one, economies of scale exist. If the value is equal to one, there are constant returns to scale. If it is greater than one, there are diseconomies of scale.

For two-year public college libraries, the sum of the b_1 's is 1.3, for two-year private college libraries it is 0.8, for four-year public it is 1.2, and for four-year private it is 1.1. Three out of four categories of libraries have diseconomies of scale. Only in two-year private college libraries are there economies of scale. Again it should be noted that the R^2 values for two-year private college libraries are relatively low and thus this regression result should be treated with caution.

Thus as the amount of output produced increases, for all types of libraries except two-year public, the cost per unit of output increases.

The equations of Table VIII represent a best fit to observations comprising each library group. Each of the equations is valid only for the range of values found in the data set. That is, one cannot extrapolate from the

equations beyond the range of fitted values. This range is very large, however, and Table IX presents the maximum values of the variables. For example, the largest number of items circulated in any two-year public college library was 492,977 items at one library.

TABLE IX					
Maximum Values for Selected Cos	t and Output Variables				

	2-Year Public	2-Year Private	4-Year Public	4-Year Private
Total Operating Expenses	2,164,449	706,746	11,779,605	14,362,801
Volumes Added	27,340	29,500	225,768	203,825
Inter-Library Lending	8,065	2,986	233,028	148,825
Inter-Library Borrowing	11,764	4,530	25,777	19,666
Reference	19,584	12,000	27,506	30,890
Circulation	492,977	473,511	2,118,773	1,957,607
Hours Open Per Year	8736	8736	8736	8736

Caution is still required in inferring the limits of the regression results. It is *not* possible to conclude that if a library operated at the maximum level of all variables, there would still be diseconomies of scale. Only certain combinations of levels of variables have been tested because those are the combinations that occur in the population. Other combinations must be tested before an inference can be made, except if that combination already occurs.

The discovery of diseconomies of scale in this study is not surprising. A review of the activities that libraries perform make it obvious that as the size of library activities increase, the complexity and unit cost of performing those activities increase more rapidly. Consider cataloging, filing, acquisitions, circulation, and interlibrary lending and borrowing. It is clear that as each new unit is handled, it requires an analysis of how transactions were processed previously. As the size of the operation increases, complexity can only increase.

There are ways to break this deadlock and they involve, among other things, automation and a rationalization of the use of labor. Computer system performance will always be affected by the size of files processed, but the time required to perform the processing is minor compared to the manual alternative. Development, implementation, and operation costs of computerized systems tend to be large and these costs must be balanced against current costs to determine the desirability of automation. Nevertheless, computer systems represent a way of reversing the trend of average cost increases with size increases.

Another way to reverse the diseconomies of scale pattern is to examine the tasks that professionals and paraprofessional library workers perform

with an eye toward improving productivity. One group of workers may be more productive (i.e., produce more output per unit of time) than another for a particular task. Productivity gains could be translated into lower unit costs and provide another way of moving toward economies of scale.

CONCLUSIONS

An analysis of 1977 data from the National Center for Education Statistics has shown major differences in the output and cost characteristics of two-year public, two-year private, four-year public, and four-year private college and university libraries. Regression analysis was employed to fit a log-log equation to the cost-output for all the types of libraries. The empirical results of fitting the equation to the data show that as output increases, cost per unit of output increases more rapidly. Except for two-year private college libraries, there are diseconomies of scale in library operations. There is no point in the operating range of the libraries where this conclusion does not hold.

This result presents a serious dilemma for library administrators. It means that unless there is a major change in the way library services are performed, average costs will continue to rise more rapidly than output. One solution is to improve productivity by increasing automation of some library activities, particularly technical processing. Another solution is better rationalization of the use of labor. This includes maintaining important distinctions between professional and paraprofessional tasks.

REFERENCES

- Beazley, R. M. (1980). "Library Statistics of Colleges and Universities, 1977: Institutional Data (Libgis III/Hegis XII)." U.S. National Center for Education Statistics, Washington, DC: Department of Library Science, Indiana State University, Terre Haute, Indiana.
- Beazley, R. M. (1981). "Library Statistics of Colleges and Universities: Trends 1968-1978 Summary Data 1977." Washington, DC: U.S. National Center for Education Statistics. (PB82-127275.)
- Carpenter, R. L. (1981). College libraries: A comparative analysis in terms of the ACRL standards. College and Research Libraries 42 (No. 1), 7-18.
- Cooper, M. D. (1979). The economics of library size: A preliminary inquiry. *Library Trends 28* (No. 1), 63-78.
- Hayes, R. M. (1979). The management of library resources: The balance between capital and staff in providing services. *Library Research 1* (No. 2), 119-142.
- Kantor, P. B. (1980). "Levels of Output Related to Cost of Operation of Scientific and Technical Libraries: The Final Report of the LORCOST Libraries Project." Cleveland, OH: Case Institute of Technology, Case Western Reserve University. (PB 81-124703.)
- Kelejian, H. H., and Oates, W. E. (1974). "Introduction to Econometrics." New York: Harper and Row.

Lawrence. G. S. (1981). "Cost Functions for University Libraries: Phase I Report." Research and Analysis Branch, Library Studies and Research Division, Office of the Assistant Vice President—Library Plans and Policies, University of California Systemwide Administration, Berkeley, CA (Research Report RR-81-2.)