

## Forecasting Academic Library Growth

*Forecasting techniques developed by government and industry are being applied to various library statistics. These techniques are explained and examples of their use discussed.*

**B**UDGETING AND PLANNING PROCESSES in academic libraries are under great stress due to tight budgets, inflation, and uncertainty about the future. University administrators are stressing the need for cost justification and the necessity for short- and long-range planning. The future appears to be characterized by increasing stress because of higher prices for materials, demands for higher salaries, and growing uncertainty about availability of funding.

The need for planning assumes a greater importance in times of decreasing resources and retrenchment. The library manager, confronted with smaller budgets and higher prices on the one hand and demands for better service on the other, must devote considerable energy and effort to systematic planning. Thoughtful and informed planning decisions are imperative if libraries are to continue to function effectively as an interface between information and the student or researcher.

The planning process is a systematic and purposeful endeavor which involves the setting of goals for the future and strategies for realizing those goals. University libraries function within a dynamic environment in which teaching and research goals and methods are changing. Library planners must plan within this environment and within the framework of the university's

goals. The role of forecasting in planning is critical. Analyses of the past and present and forecasts of probable future events and situations give the planner the framework within which decisions about future courses of action are made. Good information will not necessarily guarantee good decisions or effective planning; but no information or poor information may well result in bad decisions. At the present time, when the need for accurate predictive information is so great, library managers and planners have little analytical data about the past and poor forecasts for the future.<sup>1</sup>

The purpose of this paper is to discuss one type of information needed by library planners, namely, forecasts of library growth and related factors. Forecasting techniques will be discussed, and examples of their use in libraries will be reviewed.

The purpose of forecasting is to describe what is likely to occur under a given set of circumstances at some future time. The forecast provides an indication of results at a specified time in the future if conditions are not changed. The manager can change the likelihood of given results by changing goals and making decisions which will influence the course of events. Generally, there are three types of forecasting techniques used in libraries: qualitative, time series, and causal. These techniques may be used individually or in combination.

---

*Miriam A. Drake is head, Research Development Unit, Purdue University Libraries and Audio-Visual Center, West Lafayette, Indiana.*

### QUALITATIVE TECHNIQUES

Qualitative techniques are used when data are sketchy or nonexistent. They rely solely on human experience and judgment to assess the future. Sometimes they are formalized by the use of rating schemes to transfer qualitative factors into quantitative estimates in a systematic fashion. Examples of qualitative methods are the Delphi approach, "visionary forecast," and historical analogy.<sup>2</sup> It is difficult to find examples of systematic qualitative forecasting in academic libraries; however, judgmental forecasts have been used extensively because librarians do not have sufficient familiarity with quantitative techniques to use them effectively or to interpret the results.

### QUANTITATIVE TECHNIQUES

Quantitative forecasting techniques fall into two categories, time series analysis and causal methods.

#### *Time Series Analysis*

Time series analysis is a statistical technique which assumes that patterns in the past can be identified and that these patterns will be repeated in the future. Forecasts of library growth have relied primarily on time series analysis to identify trends and to determine growth rates of trends. The technique involves fitting a line or curve to the past data and projecting the line by means of its mathematical equation. The simplest trend line is a straight line in which the variable being projected increases or decreases by the same amount in each time period. The line is fitted by the method of least squares, so called because the sum of the squares of the deviations from the line is less than the sum of the squared deviations from any other straight line.<sup>3</sup> The deviations are the differences between the actual or observed values and the values produced by the straight line equation for each point in the past.

The straight line is described mathematically as

$$Y = a + bx$$

where:  $Y$  is the variable to be forecast;  $a$  is the value of  $Y$  at the  $x$  origin;  $b$  is the slope of the line or the value to be added or subtracted in each time period; and  $x$  is the value of time. The elements  $a$  and  $b$  are both constants as their values do not change; therefore, the projection is based on the same amount of growth or decline in each period.

Figure 1 illustrates a straight line projection of nonprofessional staff size in the median Association of Research Libraries library. The median library is not a specific library; rather, it represents a library which is at the midpoint on a scale in which Association of Research Libraries libraries are ranked from highest to lowest. Half the libraries rank above the median in nonprofessional staff size, and half rank below the median. The straight line shown was plotted from a mathematical equation calculated from data for the years 1962-1973. The dots on the graph show the actual or observed values for the median library. The straight line which has been drawn is that which "fits" most closely the dots representing actual values. The line has been extrapolated to 1980, giving an estimate of the number of nonprofessional staff in the median library assuming that the trends of 1962-1973 remain unchanged.

In reality, library growth indicators such as volumes added or volumes held do not show a straight line trend. Non-linear trends can exhibit constant growth rates or changing growth rates. A straight line shows a constant *amount* of growth, while a curve may be based on a constant annual *rate* (e.g., 20 percent) of growth or a changing rate of growth. A constant growth rate will produce an exponential curve because of the effect of compounding.

Figure 2 illustrates a projection of volumes added for the median com-

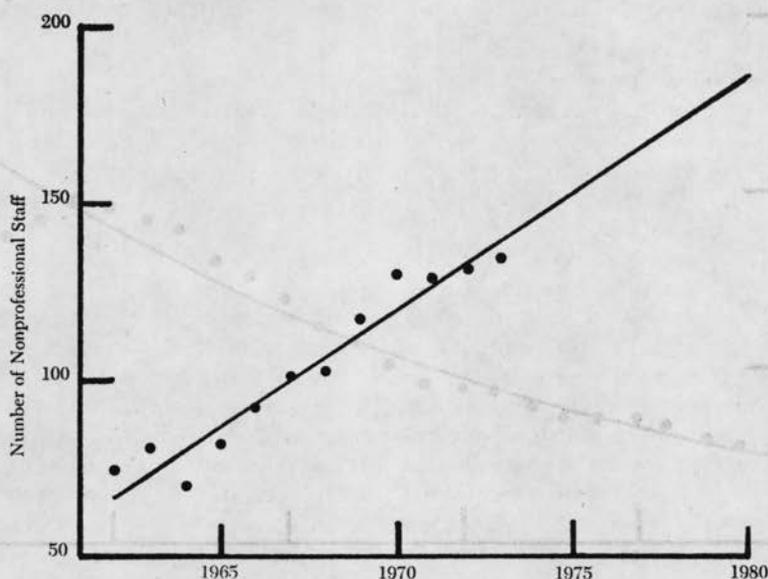


Fig. 1

Nonprofessional Staff Size in Median Library, Association of Research Libraries, 1962-1980

Source: Association of Research Libraries, *Academic Library Statistics*, 1962-1974.

posite library used in *The Past and Likely Future of 58 Research Libraries, 1951-1980: A Statistical Study of Growth and Change*.<sup>4</sup> The dots represent observed values for the median library. The curve illustrates a projection based on a constant growth rate of 6.52 percent for the years 1951-1973.

The curve illustrated in Figure 3 was also taken from *The Past and Likely Future of 58 Research Libraries, 1951-1980: A Statistical Study of Growth and Change* and shows growth at an increasing rate; that is, a percentage growth which is increasing each year. It is clear that neither curve is suitable for a short-range forecast of volumes added for the median library. Observed values are declining, and qualitative judgment regarding library budgets, inflation, etc., would indicate that these values will decline further. In this case, the forecaster must raise questions regarding both the long-range and short-run significance of the decline. How long will it contin-

ue? Does it represent a major change in pattern? Will long-run growth at previous rates resume? This case clearly illustrates the importance of experience and judgment in forecasting. The numbers alone do not tell the story.

Time series analysis, because it assumes that the future will repeat the past, "is more likely to be correct over the short term than it is over the long term, and for this reason these techniques provide us with reasonably accurate forecasts for the immediate future but do quite poorly further into the future."<sup>5</sup> The validity of that statement can be shown by looking at the forecasts shown in Figures 2 and 3. Another shortcoming of time series analysis for libraries is that it cannot predict irregular change in the rate of growth. It can predict only on past rates. Figure 2 clearly shows that actual values for 1972 and 1973 are deviating substantially from the trend line.

These deviations indicate that growth

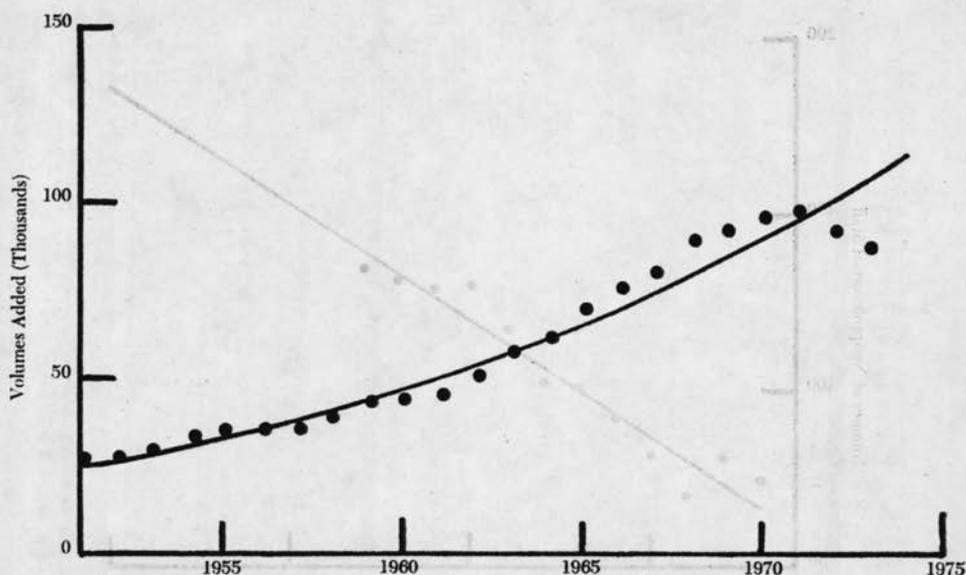


Fig. 2

Volumes Added in Median Composite Library (Constant Growth Rate), 1951-1975

Sources: O. C. Dunn, D. L. Tolliver, and M. A. Drake, *The Past and Likely Future of 58 Research Libraries, 1951-1980: A Statistical Study of Growth and Change*, 9th ed. (West Lafayette, Ind.: Purdue University Libraries and Audio-Visual Center, 1973), p.43; and O. C. Dunn, W. F. Seibert, and J. A. Scheuneman, *The Past and Likely Future of 58 Research Libraries, 1951-1980: A Statistical Study of Growth and Change*, 5th ed. (West Lafayette, Ind.: Purdue University Libraries and Audio-Visual Center, 1969), p.43.

may be changing. If the forecaster knows that there are factors which will change the rate of trend, such as budget constraints, then other methods or combinations of methods must be used to forecast the particular variables in question.

### Causal Models

The most sophisticated forecasting techniques are called causal models because they relate the values of one variable to two or more other variables. Economists use a variety of these models in different applications, including input-output, multiple regression, and econometric models. Only multiple regression models will be considered here. The multiple regression technique assumes that a dependent variable, the variable to be forecast (e.g., volumes added, total operating expenditures, etc.), is related to two or more independent or causal variables which are

assumed to be "exogenous," that is, outside the control of the dependent variable. The variables which are defined as independent are those for which values are known. These values generally will not be affected directly by changes in the dependent variable. The variable to be forecast is the dependent or "resultant" variable. Its value is related to and may be estimated from changes in the causal variables.<sup>6</sup> Mathematically, a functional relationship is assumed.

The standard equation is

$$Y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$$

where:  $Y$  is the dependent variable;  $a$  is a constant;  $x_1, x_2, x_3, \dots, x_m$  represent independent variables; and  $b_1, b_2, b_3, \dots, b_m$  represent net regression coefficients, i.e., the effect on  $Y$  of a change in  $x$  when "allowance has been made for other independent variables."

The choice of independent variables will depend on the variable to be fore-

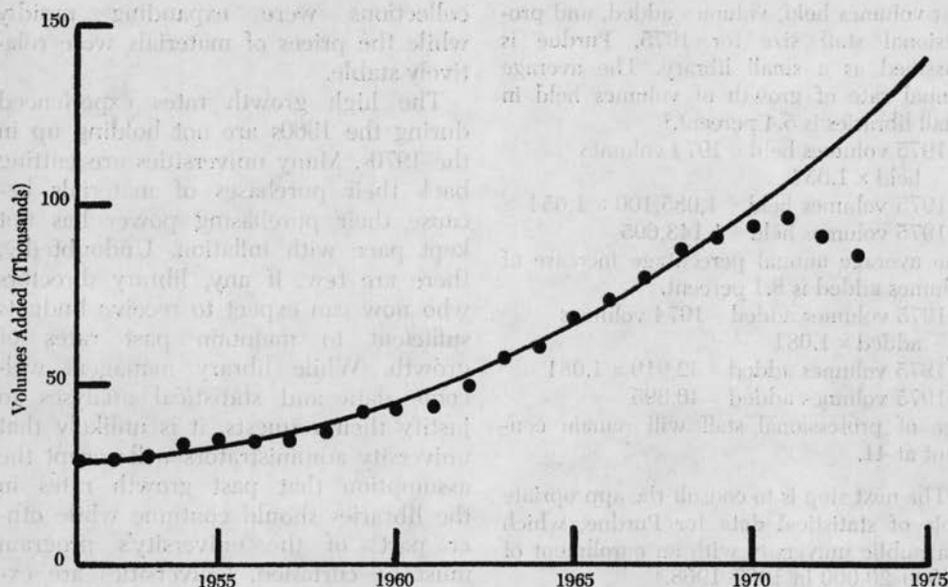


Fig. 3

Volumes Added in Median Composite Library (Increasing Growth Rate)

Source: O. C. Dunn, D. L. Tolliver, and M. A. Drake, *The Past and Likely Future of 58 Research Libraries, 1951-1980: A Statistical Study of Growth and Change*, 9th ed. (West Lafayette, Ind.: Purdue University Libraries and Audio-Visual Center, 1973), p.21, 43.

cast and the reasons for the forecast. If one were trying to forecast the probable size of the library budget in a particular institution, one would need to select variables which are not affected by the library budget, such as the university budget, Ph.D. fields, etc. Relationship or lack of relationship between these independent variables or causal variables and the library budget for some years in the past would be calculated. A forecast of the library budget can be derived based on mathematical relationships between the independent variables and the library budget and forecasts of the values of the independent variables. The forecasts of independent variables may be based on time series analysis or other techniques. The forecaster would establish the mathematical relationships between the variables and test this relationship for statistical significance, which would indicate the reliability of the forecast and the likelihood of error.

This paper will not consider tests of

significance. The interested reader can find information and formulas for these tests in any standard statistics textbook.

Baumol and Marcus in their book *Economics of Academic Libraries* combined time series analysis and multiple regression analysis to produce a method for forecasting academic library budgetary and staffing needs. Their model assumed that academic libraries would continue to grow at past rates and that the relationship of staff to materials would remain constant over time. By combining growth trends from the time series analysis and coefficients (or relationships) from the multiple regression analysis a projected library budget may be derived.

The example which follows applies the results of Baumol and Marcus analysis to a projection of total library costs for Purdue University Libraries for 1975.

The first step in the procedure is to fore-

cast volumes held, volumes added, and professional staff size for 1975. Purdue is classified as a small library. The average annual rate of growth of volumes held in small libraries is 5.4 percent.<sup>7</sup>

1975 volumes held = 1974 volumes held  $\times$  1.054

1975 volumes held = 1,085,100  $\times$  1.054

1975 volumes held = 1,143,695

The average annual percentage increase of volumes added is 8.1 percent.

1975 volumes added = 1974 volumes added  $\times$  1.081

1975 volumes added = 42,919  $\times$  1.081

1975 volumes added = 46,395

Size of professional staff will remain constant at 41.

The next step is to consult the appropriate table of statistical data for Purdue, which is a public university with an enrollment of 10,000-20,000 in 1967-1968.<sup>8</sup>

The formula for determining total operating costs is

$$Y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$$

Total operating costs, 1975 = 0.5571  
 + .0139 (volumes held)  
 + .8139 (volumes added)  
 + .1317 (professional staff)

Total operating costs, 1975 = 0.5571  
 + .0139(114.4)  
 + .8139(4.6)  
 + .1317(41)

Total operating costs, 1975 = 11.2909

Costs are expressed in units of \$100,000, so that projected operating costs for the Purdue Libraries are \$1,129,090. The actual operating costs for the Purdue Libraries in 1975 will be more than double the projected figure.

The example above illustrates how time series analysis can be applied to project values needed as input into a multiple regression analysis. The statistical techniques used by Baumol and Marcus are useful forecasting tools; however, the results of their analyses are of limited value because the data on which the analyses were based are obsolete and not reflective of the current situation. During the period analyzed by Baumol and Marcus, 1951 through 1968, library

collections were expanding rapidly while the prices of materials were relatively stable.

The high growth rates experienced during the 1960s are not holding up in the 1970s. Many universities are cutting back their purchases of materials because their purchasing power has not kept pace with inflation. Undoubtedly, there are few, if any, library directors who now can expect to receive budgets sufficient to maintain past rates of growth. While library managers welcome data and statistical analyses to justify their requests, it is unlikely that university administrators will accept the assumption that past growth rates in the libraries should continue while other parts of the university's program must be curtailed. Universities are experiencing a period of change in which resource allocations are likely to be different from the past. Forecasting beyond two or three years is difficult; nevertheless, library managers can gain insight into the dynamics of libraries by utilizing forecasting techniques.

#### CONCLUSION

Before selecting a forecasting technique, many questions should be examined: What is the reason for the forecast, and how will it be used? Is it for budget purposes? Financial planning? Space needs? What is the time horizon? Next year? Two years? Six years? What level of detail is required? How many and what items are to be forecast? What degree of accuracy is required? Are appropriate data available? Are the variables to be forecast stable over time? How much money can be spent on forecasting?

"A manager generally assumes that when he asks a forecaster to prepare a specific projection, the request itself provides sufficient information for the forecaster to go to work and do his job. This is almost never true."<sup>9</sup> The above questions suggest that forecasting is a

job shared by decision maker and forecaster. Discussion of these questions will help insure that the forecast is appropriate to managerial needs.

The need for forecasting in the academic library field is evident. Managers of libraries which are members of the Association of Research Libraries are responsible for approximately \$300 million in annual operating expenses, in institutions with enrollments of 1.3 million students. Many librarians complain that adequate forecasting cannot be

done because the data are less than perfect. Nevertheless, techniques exist which permit the utilization of existing data to provide significant and useful indications of future library growth. Effective planning for the immediate and long-range future depends on using the data we have, such as volumes added, book prices, publishing output, enrollment projections, etc., and applying the techniques to tell us what is likely to occur under different policies and strategies.

## REFERENCES

1. Notable exceptions are William J. Baumol and Matityahu Marcus, *Economics of Academic Libraries* (Washington, D.C.: American Council on Education, 1973); and O. C. Dunn, D. L. Tolliver, and M. A. Drake, *The Past and Likely Future of 58 Research Libraries, 1951-1980: A Statistical Study of Growth and Change*, 9th ed. (West Lafayette, Ind.: Purdue University Libraries and Audio-Visual Center, 1973).
2. J. C. Chambers, S. K. Mulleik, and D. D. Smith, "How to Choose the Right Forecasting Technique," *Harvard Business Review* 49:55 (July-Aug, 1971).
3. F. E. Croxton and D. J. Cowden, *Applied General Statistics* (New York: Prentice-Hall, 1955), p.265.
4. Dunn, Tolliver, and Drake, *Past and Likely Future*, 9th ed., p.43; and O. C. Dunn, W. F. Seibert, and J. A. Scheuneman, *The Past and Likely Future of 58 Research Libraries, 1951-1980: A Statistical Study of Growth and Change*, 5th ed. (West Lafayette, Ind.: Purdue University Libraries and Audio-Visual Center, 1969), p.43.
5. Chambers, Mulleik, and Smith, "How to Choose the Right Forecasting Technique," p.50.
6. Mordecai Ezekiel and Karl A. Fox, *Methods of Correlation and Regression Analysis* (New York: Wiley, 1959), p.47-48.
7. Baumol and Marcus, *Economics of Academic Libraries*, p.8.
8. *Ibid.*, p.94.
9. Chambers, Mulleik, and Smith, "How to Choose the Right Forecasting Technique," p.46.