16. Not found. This is the catch-all category to which an item is consigned whenever it cannot be located and no reason is determined. Presumably some items have been lost or stolen, others are in transit or in use. After time has been allowed for an item to turn up a decision is made whether to attempt to replace it. The gifts and exchanges unit may be able to obtain a free copy, or it may have to be purchased.

Using Time-Series Regression to Predict Academic Library Circulations

Terrence A. Brooks

Four methods were used to forecast monthly circulation totals in 15 midwestern academic libraries. In a test of one-month forecasting accuracy, the dummy regression method, a sophisticated forecasting method for cyclical data, exhibited the smallest average error. In a test of six-month forecasting accuracy the monthly mean method, a naive forecasting method for cyclical data, exhibited the smallest average error. Straight-line predictive methods, both naive or sophisticated, had significantly greater error in both accuracy tests. A remaining research question is, Why do naive forecasting methods outperform more sophisticated forecasting methods with monthly library circulation data in long-range forecasts? It is suggested that high levels of randomness in library-output statistics inhibit the performance of sophisticated forecasting methods.

INTRODUCTION

A time series is a chronological sequence of observations on a variable.[1] An example from the field of librarianship of such a variable is circulation check-outs. Library circulation counts are commonly compiled on a daily basis and aggregated into monthly or semester reports. A series of these monthly counts is a chronological sequence of observations on the variable library circulation. Consequently, it is fair to conclude that the most typical type of statistical data libraries produce is time-series data. Library literature, however, reveals little awareness of the ways that time-series data can be used for forecasting and planning.

Time-series regression techniques are regression procedures used to predict future values of a time series. They are unique only in that they use past values of a time series to predict future values of the same time series. This paper reports the application of two types of time-series regression to the problem of forecasting academic library circulation.

FORECASTING

"In library planning and decision-making, predictions are invariably required."[2] Despite Hamburg's statement, there has not been much theoretical work or practical application of forecasting methodologies to library statistics. This is in sharp contrast to the acceptance of forecasting in other disciplines. Forecasting, or trend analysis, is considered an integral part of scientific management and rational decision making. Makridakis and Wheelwright[3] describe forecasting as a tool that permits management to shield an organization from the vagaries of chance events and become more methodical in dealing with its environment. Like bureaucracies everywhere, academic libraries need tools that will enhance planning and rational decision making.

Filley and House[4] would characterize most academic libraries today as third-

Terrence A. Brooks is assistant professor at the School of Library and Information Science, University of Iowa, Iowa City, Iowa 52242.

stage growth organizations. Large and complex, these organizations have developed beyond the early rapid growth stages identified by Filley and House and now have become institutionalized with a corps of bureaucrats who plan, organize, direct, and control. Many academic librarians are similarly charged with the tasks of planning, organizing, directing, and controlling library operations. One tool to help accomplish these managerial tasks is forecasting.

There are two forecasting studies in library literature worthy of note. The first is by Drake,[5] who considered linear regression as a predictive technique. She concluded that straight-line trend projections are not the most efficient predictors in all library situations. The reason is that library data, especially circulation data, show monthly or seasonal fluctuations. Cyclicity may be one of the reasons that forecasting techniques have had a retarded application to library statistics. Cyclicity in library-output statistics means that a variable such as monthly circulation fluctuates up and down throughout the academic year. Such cyclical data demand forecasting techniques that can model their seasonality.

The most sophisticated forecasting study in library literature to date is by Kang.[6] He forecasted the requests for interlibrary loan services received by the Illinois Research and Reference Centers from 1971 through 1978 using several methods, including methods that can model cyclical data, and found regression to be the best predictive technique. He used a weighted regression formula that gave less predictive value to older observations, and greater weight to the most recent ones. The generalizability of Kang's study is severely limited, however, due to the fact that data from only one library was used.

METHODOLOGY

The purpose of this paper is to evaluate time-series regression forecasting methods with academic monthly library circulation totals. Time-series regression is a methodology that is new to library and information science, but has been used extensively in the social sciences, business, and economic literatures.

Makridakis and Wheelwright[7] give two versions of time-series regression approaches. The first time-series regression approach uses independent variables that are past values of the time series itself.[8] An example of such an approach would be using the monthly circulation totals of several months past as the predictor of next month's circulation total. This simply means that a library's circulation time series is regressed on itself at a certain time lag. There are two caveats with this technique. First, it produces a straight prediction line and thus should suffer the same problem of poor fit that was noted by Drake. Second, it is a new application, meaning that the choice of time lag has not been studied sufficiently with academic library circulation data. Hence, the choice of any particular time lag is completely arbitrary.

The second time-series regression method uses qualitative or dummy variables.[9] In the context of multiple regression, a dummy variable is a special independent variable that can take only a limited number of values such as 1 or 0. To use dummy regression for forecasting, some monthly totals of the time series are tagged by a 1, while other months of the year are given 0s. The result is a multiple regression equation that can model the seasonal patterns of library circulation totals and should perform as a more efficient predictor than straight-line methods.

To provide benchmarks for performance comparisons two averaging methods were also used as forecasting methods. These averaging methods were used because they represent the most direct and naive approach that any academic librarian could use for forecasting. For instance, a future circulation total could simply be forecast from the average of all past values of the time series. Alternatively, a particular future monthly total could be forecast from an average of past values of that particular month.

In all, four forecasting methodologies were used with Minitab,[10] a statistical software program, and circulation data from several libraries:

1. Dummy time-series regression was

used to find an equation to predict one month and six months in advance for each library. This is a sophisticated forecasting method that can model cyclical data.

2. Lagged time-series regression was used with each library's data lagged one month and lagged six months. The decision to use a one-month time lag and a six-month time lag was arbitrary. This is a sophisticated forecasting method that makes straight-line predictions. It cannot model cyclical data.

3. A simple average was made of each library's circulation totals to provide a straight-line benchmark for comparison purposes. This is a naive straight-line forecasting technique.

4. A monthly average was computed for each library for one month and six months in advance. This provided a seasonal benchmark for comparison purposes. For instance, if January and June represent the forecasts for one and six months, then data from previous Januarys would be averaged to give a forecast for the month of January. Similarly, previous Junes would be averaged to give the June forecast. This is a naive forecasting method that can model cyclical data.

DATA

A random sample of fifteen academic libraries in the Midwest submitted monthly circulation data for analysis. The states of Illinois, Ohio, Michigan and Missouri were each represented by three academic libraries, Iowa was represented by two academic libraries, and Minnesota by one academic library. The holdings of these fifteen libraries ranged from a maximum of four million book titles down to a minimum of two hundred thousand book titles.

Ten libraries contributed time series of 60 months' duration, three libraries contributed time series of 72 months' duration, one contributed 66 months, and one contributed a time series of 53 months. The most recent six months' data for each library were set aside to provide a basis for evaluating the performance of each of the four forecasting methods. Forecasts were made with each method for each of the fifteen libraries for one month and six months in advance. Each forecasted monthly total was then compared to the actual total reported by the library and an absolute percentage error (APE) was calculated. The average of the APE values for each forecasting method (the mean absolute percentage error) was then found.

An accurate forecasting method would, relative to other methods, have a small mean absolute percentage error (MAPE) across the sample of the fifteen academic libraries. An analysis of variance was performed comparing the MAPEs to see if there was a statistically significant difference among the four forecasting methods.

RESULTS

Table 1 shows the results of forecasting one month in advance. The dummy regression method had the smallest MAPE followed by the monthly mean method. These methods are capable of modeling the seasonal patterns of academic library circulations. The two straight-line prediction methods followed with the largest MAPEs.

TABLE 1

ANALYSIS OF VARIANCE OF MEAN ABSOLUTE PERCENTAGE ERROR FOR ONE-MONTH FORECASTS

Methods	n	MAPE (%)	SD (%)
Dummy regression	15	12.22	11.08
Monthly mean	15	15.52	12.65
Lag 1 Regression	15	26.26	15.45
Simple mean	15	30.19	25.48

Analysis of Variance

Source	df	SS	MS	F	
Factor	3	3288	1096	3.74	(p=.0160)
Error	56	16391	293		
Total	59	19679			

An analysis of variance (ANOVA) test on the difference among the MAPEs of the four methods proved to be statistically significant (p=0.0160). Since the null hypothesis of no difference among the population MAPEs was rejected, a multiple comparison of the sample means was indicated. The Neuman-Keuls procedure, as outlined by Meyer[11] was used. A significant difference ($p < 0.05$) was found between the MAPEs of the dummy regression and simple mean methods. There was insufficient evidence that any other pair of means differed significantly.

Table 2 shows the results of forecasting six months in advance. Dummy regression and the monthly mean methods, the two techniques that can model the seasonal patterns of academic library circulations performed better than the straight-line methods. But the relative positions of each technique have changed: the averaging methods now outperformed regression methods in both the cyclical and straight-line cases.

An ANOVA test on the difference among these MAPEs proved to be statistically significant (p=0.0166). The Neuman-Keuls procedure showed that the monthly mean method had a significantly ($p<0.05$) lower MAPE than the other three methods.

DISCUSSION

The results of this study show the superiority of forecasting methods that can model the cyclicity of academic library statistics. In a test of one-month accuracy, the sophisticated dummy regression method was superior. In a test of six-month accuracy, the naive monthly averaging method was superior.

The outstanding unanswered question at this point is why the monthly averaging method does so well in long-run forecasting relative to the performance of more sophisticated methods. It may be due to the fact that sophisticated forecasting methods are sensitive to random fluctuations in library time-series data. Random errors in library time series such as monthly circulation totals spring from all manner of human and mechanical sources; they are akin to static interferring with a radio transmission. It is a popular theme in library literature to castigate library-output statistics for their lack of reliability and validity. Childers[12] even portrays different types of library-output statistics as having different levels of random error based on the method of collection of the statistic. It would appear that high levels of randomness are preventing sophisticated forecasting techniques from modeling library circulation data closely and accurately. When sensitive methods are used to predict the future, their forecasts are wider off the mark than less sensitive methods. The phenomenon of the success of simpler methods has been observed in other studies[13] comparing forecasting methods. The next step in researching library-output statistics would seem to be measuring the amount of randomness in library-output statistics.

TABLE 2

ANALYSIS OF VARIANCE OF MEAN ABSOLUTE PERCENTAGE ERROR FOR SIX-MONTH FORECASTS

Methods	*n*	MAPE (%)	SD (%)
Monthly mean	15	12.38	10.71
Dummy regression	15	15.30	13.51
Simple mean	15	39.16	38.45
Lag 6 Regression	15	39.63	42.08

Analysis of Variance

Source	df	SS	MS	F	
Factor	3	9864	3288	3.71	(p=.0166)
Error	56	49652	887		
Total	59	59516			

REFERENCES

1. Bruce L. Bowerman and Richard T. O'Connell, *Time Series and Forecasting* (North Scituate, Mass.: Duxbury Press, 1979).
2. Morris Hamburg, "Statistical Methods for Library Management," in Ching-chih Chen, ed., *Quantitative Measurement and Dynamic Library Service* (Phoenix, Ariz.: Oryx, 1978), p.38.
3. Spyros Makridakis and Steven C. Wheelwright, *Forecasting Methods and Applications* (New York, N.Y.: Wiley, 1978).
4. Alan C. Filley and Robert J. House, *Managerial Process and Organizational Behavior* (Glenview, Ill.: Scott, Foresman, 1969), p.441.
5. Miriam A. Drake, "Forecasting Academic Library Growth," *College & Research Libraries* 37:53–59 (Jan. 1976).
6. Jong Hoa Kang, "Approaches to Forecasting Demands for Library Network Services" (Master's thesis, Univ. of Illinois at Urbana-Champaign, 1979).
7. Makridakis and Wheelwright, *Forecasting Methods.*
8. Ibid., p.81–84.
9. Ibid., p.217–20.
10. Thomas A. Ryan, Brian L. Joiner and Barbara F. Ryan, *Minitab Reference Manual* (University Park, Pa.: Statistics Department, The Pennsylvania State Univ., 1982).
11. Merle E. Meyer, *A Statistical Analysis of Behavior* (Belmont, Calif.: Wadsworth, 1976).
12. Thomas A. Childers, "Statistics that Describe Libraries and Library Service," in Melvin J. Voigt, ed., *Advances in Librarianship 5* (New York, N.Y.: Academic Press, 1975).
13. S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition," *Journal of Forecasting* 1:111–53 (Apr.–June 1982).