INTERFACES
Vol. 9, No. 4, August 1979

0092-2102/79/0904/0049$01.25

BOX-JENKINS VS MULTIPLE REGRESSION: SOME ADVENTURES IN FORECASTING THE DEMAND FOR BLOOD TESTS

Everette S. Gardner, Jr.*
Graduate School of Business Administration, University of North Carolina,
Chapel Hill, North Carolina 27514

ABSTRACT. This paper reports the application of a multiple regression forecasting model in a hospital laboratory setting. Twenty-five time series models were tested, and the comparative results should be of interest to practitioners. Box-Jenkins theory could not be stretched to fit the time series, which is characterized by powerful but erratic trend and seasonal components. Simple exponential smoothing performed much better than several highly sophisticated smoothing models.

Introduction

Quantitative forecasting models are still rare in health care management, despite a wide range of opportunities for their application. One opportunity was found at the North Carolina Memorial Hospital (NCMH), where forecasting the demand for laboratory services has become an important function in planning and budgeting. Laboratories at NCMH generated more than $9.5 million in revenues in the fiscal year ending June 30, 1978, although no formal models were used to project most of the revenues.

This paper describes the implementation of a multiple regression forecasting model for the Clinical Coagulation Laboratory at NCMH. The model forecasts the aggregate number of laboratory tests to be conducted each month as a function of time and incorporates dummy variables for seasonality. Aggregate forecasts are broken down into individual types of tests using simple exponential smoothing models, in order to forecast the percentages of the aggregate accounted for by each type. Short range forecasts are used in manpower planning and inventory control, while long range (18 month) forecasts are the basis for budgeting and revenue projections for the state government.

The multiple regression model was implemented after an evaluation of some 25 time series models, detailed in the Appendix. The comparative performances of the models tested were surprising and should be of interest to model builders and managers faced with similar data.

The forecasting problem

The annual budget cycle at the NCMH begins in early January. Data through December must be used to project monthly tests beginning in June and running through July of the following year. In January, 1979, for example, we must project monthly tests

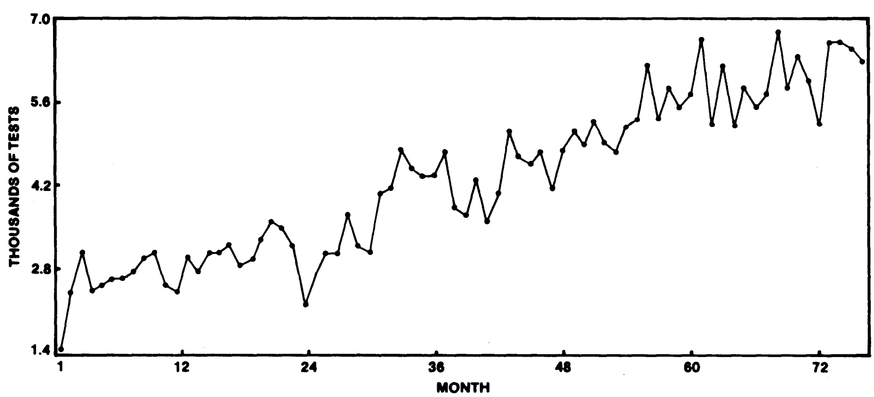
*Present address: LCDR Everette S. Gardner, Jr., SC, USN, USS Dwight D. Eisenhower (CVN69), FPO New York, New York 09501.

HEALTH CARE — HOSPITALS: FORECASTING

from July, 1979, through June, 1980. Short range forecasts are also needed each month using data through the previous month. Although more than 20 individual types of test are conducted in the Laboratory, three types usually make up more than 85% of the volume. To provide an acceptable level of detail for budgeting and inventory control, both long and short range forecasts must include individual estimates for the three major types plus an "all others" category. Fortunately, the percentages of aggregate demand for the major types have been fairly stable over time.

A plot of aggregate monthly tests for the last six years in the Laboratory is shown in Figure 1. A strong but highly erratic trend has increased the level of the series by about 475% from period 1 to the peak in period 68. Some of the variability in the series can be explained by an interesting seasonal pattern. Demand usually falls off in December due to the reluctance of potential patients to seek medical care during the holiday season. Demand picks up in January and then gradually declines until August, when a strong spike in demand occurs as doctors are rotated in the hospital and become familiar with new jobs. This "training effect" in demand usually continues through October. Unhappily for the model builder, the seasonal pattern is erratic. A close examination of the plot shows that the peak and trough months as well as the magnitude of seasonal fluctuations vary considerably from year to year. This sort of behavior makes it difficult to predict whether seasonal parameters would be useful in developing a forecasting model.

FIGURE 1. Total patient tests, Clinical Coagulation Laboratory, North Carolina Memorial Hospital, monthly, January, 1972 to April, 1978.



Model development

Simplicity is always cheering to model users, and in this case it was indispensable. Budget constraints were such that any model would have to be routinely maintained by the Laboratory staff, who have modest statistical backgrounds. An annual checkup on the model by a consultant would be possible, but no more than that. Such considerations rule out most Box-Jenkins models. Perhaps the major obstacle to Box-Jenkins forecasting is that the models are often unintelligible to managers, which is always dangerous in real applications.

Box-Jenkins models were tested, however, along with exponential smoothing and multiple regression models, on the off chance that something simple might apply. A detailed evaluation of the results is given in the Appendix. In general, the regression models gave significantly better forecast errors for both short and long range horizons. For example, in one-period-ahead forecasting tests, the group of multiple regression models tested gave mean absolute forecast errors in the range of 5.0 to 5.9%, compared to 7.7% for the best Box-Jenkins model, and 9.4% for exponential smoothing. These results led to the selection of the simplest multiple regression model, which relies only on time and seasonal dummy variables.

Implementation

The only major implementation problem was in establishing an accurate data base for the model. The first version of the time series in Figure 1 was much more variable than the final version shown and yielded mean squared errors for most models from 25 to 30% higher than the final values given in the Appendix. The author requested that a number of outlying data points be validated in the first series and the results uncovered inconsistencies in bookkeeping procedures throughout the data. About one man-week was expended in reconstructing all components of the aggregate series in Figure 1. This exercise proved to be quite useful in establishing bookkeeping standards in the Laboratory.

The multiple regression model was originally developed in late 1977 and was used by the Laboratory Director in January, 1978, to develop forecasts for the fiscal year beginning in July, 1978. Short-range forecasts are being used to plan vacation schedules for the 27 employees of the Laboratory and to order supplies for the tests. The model is set up to run in the Time Series Processor (TSP) computer package [6]. Simple exponential smoothing models are used to forecast percentages for the major tests, which are then applied to the aggregate forecasts from the regression model. The Laboratory Director's secretary types several cards each month to record new data and the model generates listings and plots of forecasts for the next 18 months.

Dollar savings from the model implementation are impossible to estimate. Records of on-hand inventory values were not kept prior to model implementation, so we cannot prove any inventory savings. It would also be difficult to attach a specific dollar value to having a rational basis for budget preparation in the Laboratory. We can make an interesting comparison, however, to an intuitive forecast made by the Laboratory in January, 1977. This forecast for total tests from July, 1977 through June, 1978 was about 20% high, which caused the Hospital to anticipate approximately $100,000 in revenues that did not materialize. The multiple regression model, using data through December, 1976, has a total error of less than 1% for the same period. A similar comparison in November, 1977, was the basis for the Laboratory Director's decision to implement the model.

Results

Forecasting performance over the last eight months has been excellent. Forecast errors for the three major types of tests have averaged 4.4%. Twenty of the 24 recorded errors have been less than 5.5%. (The outlying errors were 6.1, 6.3, 7.5, and 8.3%.)

Conclusion

As previously stated, quantitative forecasting models are rare in health care management. This paper demonstrates that the demand for blood tests is reasonably predictable, despite its extreme variability.

REFERENCES

[1] Box, G. E. P., and Jenkins, G. M., *Time Series Analysis, Forecasting and Control*, Holden Day, Inc., San Francisco, 1970.

[2] Cochrane, D., and Orcutt, G. H., "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 32—61.

[3] Dancer, R. E., and Gray, C. F., "An Empirical Evaluation of Constant and Adaptive Computer Forecasting Models For Inventory Control," *Decision Sciences*, Vol. 8, No. 1 (Jan. 1977), pp. 228—238.

[4] Gilchrist, W. *Statistical Forecasting*, John Wiley and Sons, New York, 1976.

[5] Groff, G. K., "Empirical Comparison of Models for Short Range Forecasting," *Management Science*, Vol. 20, No. 1 (Sept. 1973), pp. 22—31.

[6] Hall, B. H., and Wyse, L. E., *Times Series Processor User's Manual*, Triangle Universities Computation Center, Research Triangle Park, N. C., 1977.

[7] Johnston, J., *Econometric Methods*, 2nd Ed. McGraw-Hill, New York, 1972.

[8] Mckeown and Lorek, "A Comparative Analysis of the Predictive Ability of Adaptive Forecasting, Reestimation, and Reidentification Using Box-Jenkins Times Series Analysis," working paper #327, University of Illinois, Urbana-Champaign, August, 1976.

[9] Nelson, C. R., *Applied Time Series Analysis*, Holden-Day, Inc., San Francisco, 1973.

[10] Starr, M. K., *Management: A Modern Approach*, Harcourt Brace Jovanovich, Inc., New York, 1971.

[11] Trigg, D. W., and Leach, A. G., "Exponential Smoothing with an Adaptive Response Rate," *Operational Research Quarterly*, Vol. 18, No. 1 (1967), pp. 53—59.

[12] Winters, P. R., "Forecasting Sales by Exponentially Weighted Moving Averages," *Management Science*, Vol. 6, No. 3 (April. 1960), pp. 324—342.

Appendix — evaluation of forecasting models

A summary of the mean squared error (MSE) results for models fitted on all 76 data points in Figure 1 is given in Table 1. The MSE index is the ratio of the MSE of each model to Model 24. The model implemented was Model 22. Although its fitted MSE over all the data is higher than the last three models, it yields about the same forecasting errors over the last 16 periods of the data.

The Box-Jenkins models were tested for the short range application, but gave disappointing results. The models listed are believed to exhaust the possibilities in Box-Jenkins theory [1], [9] for this data. In the author's opinion, the poor showing of the Box-Jenkins models can be attributed to the emphasis on parsimony in the theory. For seasonal data, Box-Jenkins theory requires that every month be highly correlated with the same month one year back. This gives a strong autocorrelation coefficient between data points separated by 12 months, and can usually be modeled with a single moving average parameter for seasonality. In this data, most months do not show consistent seasonal autocorrelation, which confounds the theory. As Starr [10] observes, parsimonious theory often does not fit the messy state of affairs typically encountered in real applications.

It is interesting to contrast these results with Groff's research [5]. Groff found that many standard Box-Jenkins models yielded higher forecasting errors than a variety of other models over a wide range of data. In Groff's research, however, there was no attempt to fit all possible Box-Jenkins models justified by the data; here that attempt was faithfully made, and the regression models gave superior performance.

It would be foolish to generalize from these comparisons that the expense required to fit Box-Jenkins models cannot be justified by improved forecasting in other applications. However, in the author's opinion, the applied model builder should save Box-Jenkins theory as a last resort. Particularly in seasonal data, the theory is tedious and equivocal as to which models might apply. Frequently, one must test several models for a single time series, at a cost of 1-2 minutes of CPU time each, depending on the program used. Given an acceptable Box-Jenkins model, then there is no assurance that the same model will give the best results after a few more periods of data are recorded. Mckeown and Lorek [8], for example, report an application in which the best-fitting Box-Jenkins model changed three times in ten periods on a single time series. This is probably an extreme case, but it emphasizes the severe maintenance problems inherent in Box-Jenkins forecasting. Maintenance of other types of models is always much easier and less expensive than Box-Jenkins.

The exponential smoothing models tested were also rather miserable performers. Notice in Table 1 (Models 12-16) that smoothing an erratic trend and/or an erratic seasonal pattern was worse than simply ignoring these factors. (To avoid biasing the results by a poor choice of smoothing constants, all constants from 0.0 to 1.0 in steps of 0.05 were enumerated.) The series was deseasonalized and the experiment was repeated with the applicable Models 17-20. Again, simple smoothing was superior. The author has obtained similar results in several other applications, which suggests that practitioners using the more sophisticated smoothing models might do well to compare the performance of a simple model with a relatively high smoothing constant.

Adaptive smoothing approaches, such as some variation of the Trigg and Leach model [11], were considered but not tested. As Trigg and Leach point out, adaptive models tend to overreact to large fluctuations in the data which are not part of step or ramp increases in the level of the series. The wild behavior of the present series over the last 16 periods makes adaptive smoothing seem unpromising.

The strategy in developing the regression models listed in Table 1 was to start with the simplest possible model, the simple linear regression on time, and then to progressively enrich the model with (1) seasonal dummy variables, (2) lagged values of laboratory tests as explanatory variables, and (3) adjustments for autocorrelation in the residuals of the models using the Cochrane-Orcutt technique [2], [7]. These refinements gave a MSE in Model 24 that was 43% better than the best Box-Jenkins model and 40% better than simple exponential smoothing on deseasonalized data.

TABLE 1. Mean squared error (MSE) summary.

Box-Jenkins Models [1], [9]

Differencing	Moving Average Parameters at Lags	Autoregressive Parameters at Lags	MSE Index
1. First	1, 4, 5, 12	—	146
2. First	1, 4, 5	—	143
3. First	1, 4		150
4. Second	—	1, 2	245
5. Second	1	1	229
6. Seasonal	1 x 12	—	165
7. Seasonal	1, 12	—	168
8. Seasonal	1, 12	12	152
9. Seasonal	1, 12	—	176
10. First and Seasonal	1	—	215
11. Second and Seasonal	1	—	354

Exponential Smoothing Models — Raw Data

12. Simple Smoothing, alpha = 0.5	169
13. Double Smoothing [4]	242
14. Triple Smoothing [4]	331
15. Holt's Linear Trend Model [4]	254
16. Winters' Linear Trend and Seasonal Model [11]	216

Exponential Smoothing Models — Seasonally Adjusted Data

17. Simple Smoothing, alpha = 0.5	140
18. Double Smoothing	209
19. Triple Smoothing	299
20. Holt's Linear Trend Model	211

Regression Models

21. Simple Linear Regression on Time	137
22. Multiple Linear Regression on Time and Monthly Dummy Variables	120
23. Same as 22 with Independent Variable Lagged One Period	104
24. Same as 22 with Independent Variable Lagged Both One and Two Periods	100
25. Cochrane-Orcutt Regression on Time and Monthly Dummy Variables	107