

A Simulation Model for Purchasing Duplicate Copies in a Library

W. Y. ARMS: The Open University, and T. P. WALTER: Unilever Limited. At the time this study was undertaken the authors were at the University of Sussex.

Provision of duplicate copies in a library requires knowledge of the demand for each title. Since direct measurement of demand is difficult a simulation model has been developed to estimate the demand for a book from the number of times it has been loaned and hence to determine the number of copies required. Special attention has been given to accurate calibration of the model.

INTRODUCTION

A common difficulty in library management is deciding when to buy duplicate copies of a given book and how many copies to buy. A typical research library has several hundred thousand different works; many are lightly used but all are potential candidates for duplication. The problem which we faced at Sussex University was how to obtain reliable forecasts of the demand for each title and to translate this into a purchasing policy. At present Sussex spends between £10,000 and £20,000 (\$22,000-\$44,000) per year on duplicate copies, and as the university grows this amount is increasing steadily.

Because of the large number of books in a library relatively little data are available about each title. Records are kept of books on loan or removed from the library, but frequently these are the only routine data collected. Few large libraries even manage inventory checks. We therefore looked for a system that could be implemented with the minimum of data collection, preferably one based on existing records.

FORECASTS OF DEMAND

If the demand for a particular book is known, it is possible, though not necessarily easy, to determine how many copies of that book are needed to achieve a specified level of service, such as a copy being available on 80 percent of the occasions that a reader requires the book. Unfortunately demand cannot be measured directly, even retrospectively. Records of the

number of times that a book is issued from the library contain no information about how many times the book was used within the library, nor how many readers failed to find a copy and went away unsatisfied. Since both these factors are extremely difficult to measure, one of the central parts of our work was to develop a method of estimating them from data readily available.

To forecast demand two lines of approach seemed reasonable: subjective estimation based on faculty reading lists; and forecasts based on the number of loans in previous years. In the past, Sussex Library has made extensive use of reading lists provided by faculty to decide how many copies to buy of each title. As the books most in demand are those recommended for undergraduate courses this seemed a sensible approach, though the number of copies required is not obvious even if the demand is known. Webster analysed the effectiveness of these lists in predicting demand for specific titles and evaluated the purchasing rule being used, one copy for every ten students taking a course.¹ Restricting his attention to books known to be in demand and marked in the catalog, he drew a random sample of 673 titles, about 4 percent of the books falling into this category. He compared the number of loans of each of these titles over a term with data from the reading lists supplied at the beginning of the term. As the library had made a special effort to obtain reading lists for all courses taught that term, he had data on the number and type of students taking each course, the importance given to each text, and the subject areas involved. Yet despite a thorough analysis of these data Webster was able to find very little relationship between observed demand and reading list information. His work shows that faculty at the university have remarkably little knowledge of the books that their students read. In the sample some books strongly recommended to large groups of students were hardly used and some of the most heavily used works appeared on no reading list. The results of this study are fascinating from an educational viewpoint but less satisfying as operational research.

The failure of this approach led us to predicting demand from records of the number of past loans. This divides into two parts: using the number of loans over a period to estimate what the total demand was during that period; and using this estimate of the demand in one period to forecast the demand in another. Various evidence suggests that the latter is a sensible thing to do. The main demand for heavily used books comes from undergraduate courses. Most faculty are loyal in their reading habits, recommending books they know rather than new ones, and each course tends to be repeated year after year with a syllabus that changes only gradually. The use of past circulation to forecast future use is fundamental to a Markov model of book usage developed by Morse and Elston and tested with data from the M.I.T. Engineering Library.² For our work we have used the number of loans in a given term to predict the demand in the corresponding term a year later.

Estimating the total demand in a period from the number of loans in that period is more difficult. This requires a model of the circulation system.

MATHEMATICAL APPROACH

Several attempts have been made to apply the methods of inventory control or queueing theory to the problem of buying duplicates. For example, Grant has recently described an operational system using the simple rule that the number of copies required to satisfy 95 percent of the demand is

$$n(\mu_s + 2\sigma_s)/t$$

where n is the number of times that the book is issued during a period of t days and μ_s and σ_s are the mean and standard deviation of the time that each book is off the shelf when on loan.³

This type of approach has the advantage of being straightforward to use. Periodically a simple computer program analyzes the circulation history of each book in the library and prints a list of books requiring duplication. However, the method suffers from difficulties both mathematical and practical. To obtain the simple mathematical expression given above, several simplifying assumptions have to be made. For example, the expression ignores use of a book within the library, and identifies demand in a period with the number of loans within that period. Practical difficulties in arriving at a more exact mathematical expression are discussed in the next section.

DIFFICULTIES IN CONSTRUCTING A MODEL

The following are the main difficulties that we found in constructing a model, either mathematical or using simulation:

1. The most useful measure of the effectiveness of a duplication policy is *satisfaction level*, the proportion of readers who on approaching the shelves find a copy of the book there, but satisfaction level is almost impossible to measure directly since, although some unsatisfied readers ask that the book be held for them, most go away without comment. More or less equivalent is the *percentage time on shelf*, the proportion of time that at least one copy of the book is available. This can be measured directly, though a visit to the shelves is needed, and was found useful in validating our model. If the underlying demand is random these two measures of effectiveness have the same value.
2. Use of books within the library is also difficult to measure. At Sussex, as in most libraries, data are available only on the number of times that a book is lent out of the library. If a reader does not find a copy on the shelves or if he uses a book within the library but does not take it away then no record is generated. Since various studies, notably that of Fussler and Simon, suggest that the amount of use within li-

braries often exceeds the number of loans recorded by a factor of three or more, if the number of loans is used to estimate demand a reasonable knowledge of within-library use is essential.⁴

3. The number of copies required to achieve a specified satisfaction level does not go up linearly with demand. Since a reader is satisfied if he finds a single copy on the shelves, proportionately fewer duplicates are needed of the books most in demand. At Sussex more than twenty copies are provided of several books and this nonlinearity is very noticeable.
4. The demand for a title is erratic, changing from term to term, from week to week, and from day to day, even if the mean demand is constant. Over a period such as a term three different effects might be expected: a background random demand independent of university courses; sudden peaks when a book is required for a course taken by several students; and feedback caused by previously unsatisfied readers returning.
5. The circulation of books is surprisingly complicated. At Sussex some books are designated short term loan and can be borrowed for up to four days only; the remainder are long term loan books and can be borrowed for up to six weeks. Circulation data show that the time for which a book is off the shelf is not the same as the period for which it is lent, but has a heavily skewed distribution. Few books are returned until near the due date; just before the book is due back there is a peak when most books are returned but many become overdue and the tail of the distribution dies away slowly.

SIMULATION

As these various factors seemed too complex to derive usable mathematical results, we decided to use computer simulation of the book circulation. Simulation of book circulation is not new. In particular it has been used at Lancaster University by Mackenzie et al. to decide loan periods.⁵ Their report includes a good description of the general approach.

The object of our simulation was to model the circulation process so that we could study the relationship between three groups of parameters:

1. *Observed data*
Number of copies available
Number of loans
2. *Total underlying demand*
3. *Measures of effectiveness*
Satisfaction of level
Percentage time on shelf.

The results obtained from any simulation are only as accurate as the values given to the variables used to calibrate the model. As several of these values were not known at all accurately when the work was begun, special efforts were put into careful validation and calibration of the mod-

- braries often exceeds the number of loans recorded by a factor of three or more, if the number of loans is used to estimate demand a reasonable knowledge of within-library use is essential.⁴
3. The number of copies required to achieve a specified satisfaction level does not go up linearly with demand. Since a reader is satisfied if he finds a single copy on the shelves, proportionately fewer duplicates are needed of the books most in demand. At Sussex more than twenty copies are provided of several books and this nonlinearity is very noticeable.
 4. The demand for a title is erratic, changing from term to term, from week to week, and from day to day, even if the mean demand is constant. Over a period such as a term three different effects might be expected: a background random demand independent of university courses; sudden peaks when a book is required for a course taken by several students; and feedback caused by previously unsatisfied readers returning.
 5. The circulation of books is surprisingly complicated. At Sussex some books are designated short term loan and can be borrowed for up to four days only; the remainder are long term loan books and can be borrowed for up to six weeks. Circulation data show that the time for which a book is off the shelf is not the same as the period for which it is lent, but has a heavily skewed distribution. Few books are returned until near the due date; just before the book is due back there is a peak when most books are returned but many become overdue and the tail of the distribution dies away slowly.

SIMULATION

As these various factors seemed too complex to derive usable mathematical results, we decided to use computer simulation of the book circulation. Simulation of book circulation is not new. In particular it has been used at Lancaster University by Mackenzie et al. to decide loan periods.⁵ Their report includes a good description of the general approach.

The object of our simulation was to model the circulation process so that we could study the relationship between three groups of parameters:

1. *Observed data*
 Number of copies available
 Number of loans
2. *Total underlying demand*
3. *Measures of effectiveness*
 Satisfaction of level
 Percentage time on shelf.

The results obtained from any simulation are only as accurate as the values given to the variables used to calibrate the model. As several of these values were not known at all accurately when the work was begun, special efforts were put into careful validation and calibration of the mod-

el. A separate study was made for a small sample of books, to compare the percentage time on shelf estimated by the simulation with the actual time for which a copy was available, found by looking at the shelves. The results of this study were used to check the amount of use within the library. By this means we were able to verify the simulation model and calibrate it to a highly satisfactory level of accuracy.

DESCRIPTION OF PROGRAM

The basic layout of the simulation is shown in Figure 1. This is a time advance model with a period of one day. The program has been coded in FORTRAN and running on the ICL 1904A computer at Sussex takes about one second of machine time to simulate two years. This fast speed has enabled us to try a wide range of values for most parameters and to experiment with a variety of distributions of arrival times and book return dates.

1. *Satisfaction level*

At the beginning of each day the number of demands for that day is generated. The satisfaction level is taken as the proportion of these requests which can be satisfied from the books left on the shelf from the previous day and those returned during the simulated day.

2. *Within-library use*

The proportion of use that takes place within the library was a key parameter in calibrating the model. The first version of the simulation program assumed a figure of 25 percent use within the library. This was based on a small survey of the type of books being studied, standard texts used for undergraduate courses. The weakness of this survey was that it used a count of those books that were left lying in the library at the end of the day and did not make sufficient allowance for books reshelved by readers or by library staff during the day.

The validation experiment showed a consistent difference between predicted and observed percentage time on shelf which could be corrected by changing the value of the within-library use parameter to 60 percent.

3. *Distribution of demand*

Two distributions of demand have been used, Poisson arrivals with a specified mean, and a step demand superimposed on a Poisson process. In both cases provision is made for a proportion of unsatisfied readers to return later. As the effect of this feedback is to introduce sharp peaks of demand, the two distributions have proved surprisingly similar in the results produced and most of the runs of the program have been done with random demand.

A recent survey showed that 69 percent of readers who fail to find a book intend to return, but we do not know how many actually come back nor what the time interval is before they return.⁶ The simulation proved to be insensitive to moderate changes of these parameters

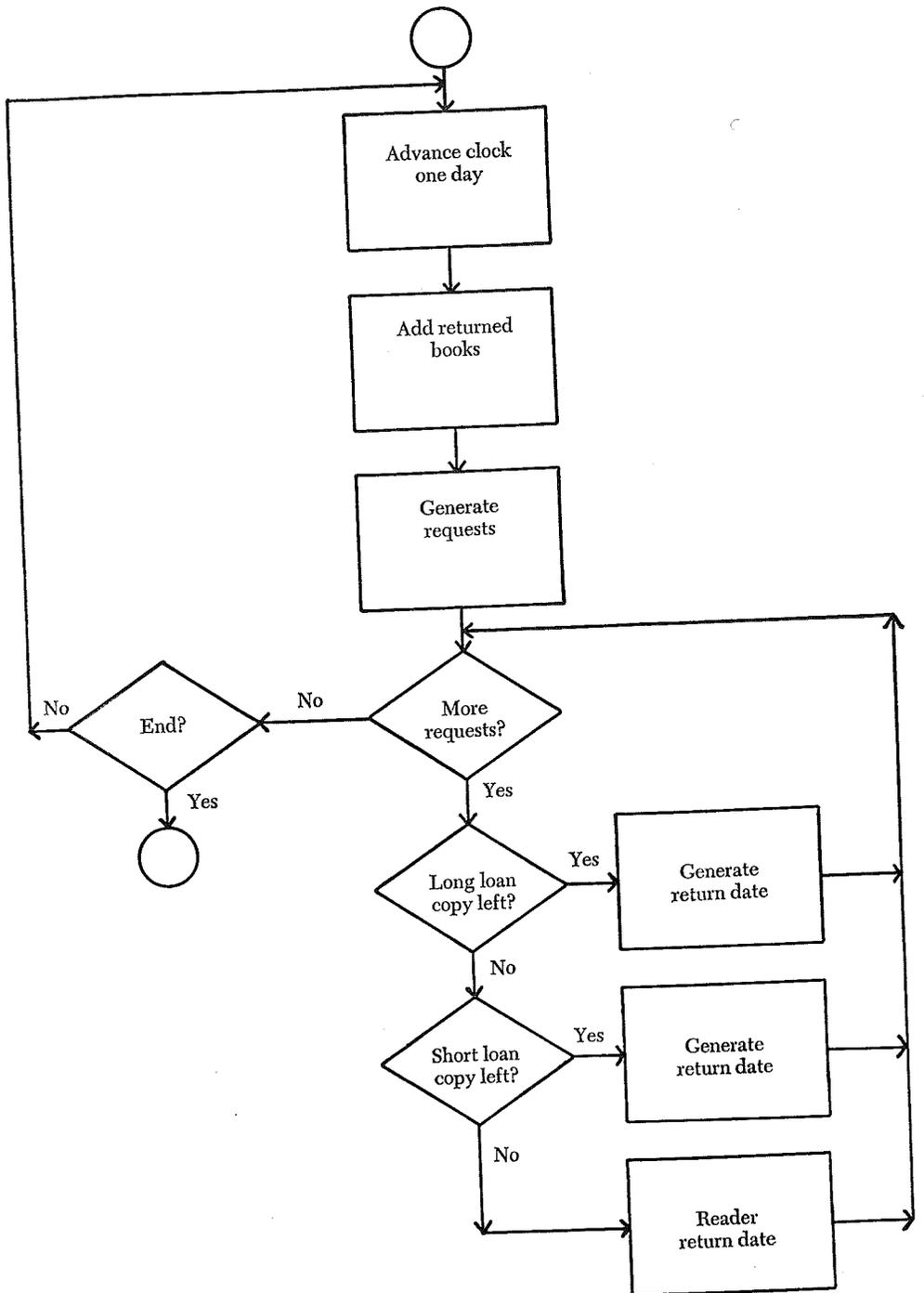


Fig. 1. Outline flowchart of simulation program

and for most runs 25 percent of unsatisfied readers were deemed to return after a delay which averaged two days.

4. *Period for which the book is off the shelf*

The simulation allows for a book to be borrowed within the library, in which case it is available again the next day, or to be lent from the library. If the book is lent, the return date is generated from one of two histograms which respectively refer to books available on short and long term loan. These histograms were derived from an analysis of all books returned during one week in autumn 1970, modified to reflect changes in the circulation system.

VALIDATION EXPERIMENT

Although the structure of the simulation is fairly straightforward several parameters used in the model have been estimated indirectly. Validation of the model took two forms. Firstly we ran the program with a wide range of values for the main parameters to see which most influence the results. Secondly a small study was set up to measure the percentage time on shelf of a number of books. For each book, the actual availability was estimated by the simulation from the number of loans during the same period.

Twenty-eight books known to be in heavy demand were selected, half in physics and half in sociology. Over a period of eight weeks the shelves were inspected once per day, at random times during the day, to see if a copy was available. The number of loans of each copy of each book during the period was noted and the library staff carried out a thorough check to determine whether any copies shown in the catalog had been lost, stolen, or had their loan category altered. The simulation was used to estimate the percentage time on shelf and this was plotted on a graph against the observed percentage.

Figure 2 shows the graph for the original values of the parameters. In this graph the x axis shows the percentage time on shelf predicted by the simulation; the y axis shows the percentage observed. If the model were perfect the points would lie near the line $y = x$, deviations being caused by y being a random variable. The graph in Figure 2 is clearly convex downwards showing a consistent error in the model, with these values of the parameters. Knowing that the simulation is sensitive to the parameter giving the proportion of use that takes place within the library and that our estimate of its value was not precise, a series of graphs were prepared varying this parameter. Figure 3 shows the same observations plotted against predictions assuming 60 percent use within the library, the value which best predicts the observations. This graph is much closer to being linear than Figure 2.

The next question is whether the nonlinearities in Figure 3 are the type to be expected from y being a random variable. A very rough calculation helps to answer this question. If we make the dubious assumption that

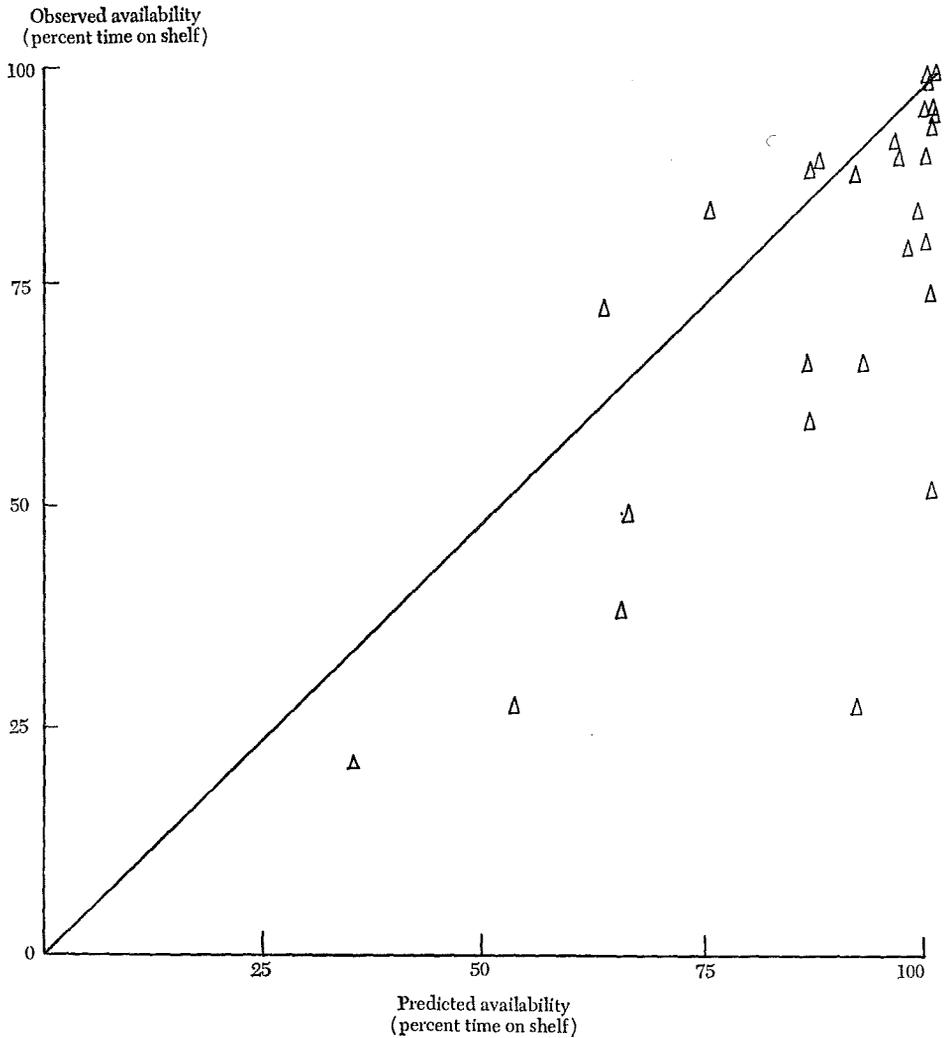


Fig. 2. Observed percentage time on shelf against predicted (25 percent use within library)

availability of a copy on a given day is independent of the days before and afterwards, then, for x given, y should be approximately normally distributed with mean x and variance $x(1 - \frac{x}{n})$, where n is the number of days in the study (forty). If this calculation were exact, 95 percent of the observations of y would lie within two standard deviations of x , but, since the assumption of independence is definitely false, we would expect the number of observations which fall within the range to be less than 95 percent.

The curves

$$y = x \pm 2 \{ x(1 - x)/n \}^{1/2}$$

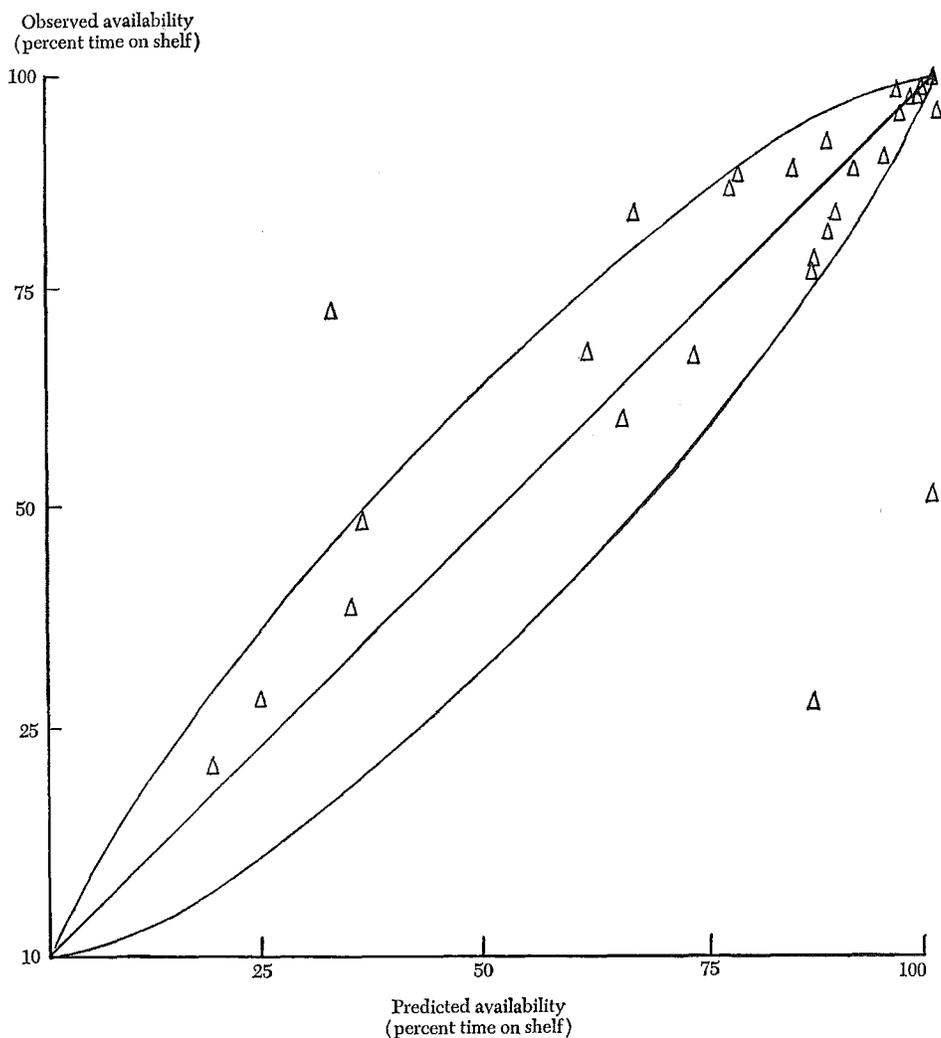


Fig. 3. Observed percentage time on shelf against predicted (60 percent use within library) with 95 percent probability curves

have been added to Figure 3. Two points lie well off all graphs and cannot be explained except as the result of books being stolen or lost during the period of the study. Of the remaining twenty-six all but three lie within the curves. This shows that the simulation model as finally calibrated gives a very reasonable description of the situation.

OPERATIONAL EXPERIENCE

The results of this simulation have been used by library staff since the middle of 1971 initially on an experimental basis. A two-stage process is in-

volved. From the computer based circulation system can be found the number of times that each short term loan copy has been circulated. From these figures the library staff can estimate the demand for a title, over a given period. Once the demand has been estimated the staff can use the simulation again to determine how many copies would have been required to have achieved a specified satisfaction level, perhaps 80 percent. If fewer copies are held by the library orders are placed for extra copies. At present these procedures are done manually using tables, but the possibility exists of modifying the computer system to identify those titles which need extra duplication. The actual decision to purchase needs to be done by library staff who can take account of factors not included in the simulation, such as price and changes of undergraduate courses.

CONCLUSION

Although this work was carried out during 1971, we shall have little operational experience of the method in action until the computer circulation system is reorganized. In the past, different copies of the same book have been processed entirely independently, meaning that the total number of loans of a given title can only be found by manually adding up the number of loans of each copy. In the revised computer system this will be done automatically. Experience will probably show that the best procedure combines use of the simulation model with reading lists and the skill of a librarian. One possible feature of a computer based system is that it could automatically indicate which books appear to require duplication.

The method used here would seem to apply equally well to other libraries. Naturally the circulation patterns of other libraries are different, which means that a different simulation would be needed, but this work has shown that it is possible to calibrate a simulation accurately enough to examine the circulation of individual books.

ACKNOWLEDGMENTS

We would like to thank the many members of the University of Sussex library staff who have helped at various stages, particularly P. T. Stone who was closely involved throughout.

REFERENCES

1. P. F. Webster, *Provision of Duplicate Copies in the University Library*, Final year project report (University of Sussex, 1971).
2. P. M. Morse and C. R. Elston, "A Probability Model for Obsolescence," *Operations Research* 17:36-47 (1969).
3. R. S. Grant, "Predicting the Need for Multiple Copies of Books," *Journal of Library Automation* 4:64-71 (June 1971).
4. H. H. Fussler and J. L. Simon, *Patterns in the Use of Books in Large Research Libraries* (Chicago: Univ. of Chicago Pr., 1969).
5. A. G. Mackenzie et al., *Systems Analysis of a University Library*. Report to OSTI on Project S1/52/02, 1969.
6. J. Urquhart, Private discussion, 1971.