TIM SAUER

Predicting Book Fund Expenditures:
A Statistical Model

A statistical model is developed to predict, at any point in the fiscal year, the amount of money that will be spent, within a book acquisition system, on firm orders against a given fund and thereby predict the number of orders that must still be placed in order to totally spend but not overspend that fund.

ONE OF THE UNRESOLVED annual problems facing many library acquisition departments is how to insure that each acquisition fund is fully spent at the end of each fiscal year.[1] The significance of this problem increases as the buying power of acquisitions budgets continues to shrink. Of the methods used to spend budgets exactly, neither underspend nor overspend by more than a few percentage points, some border on the unethical, while others require vast expenditures of staff time.

In some of the more fortunate academic libraries, the problem only exists for the total acquisitions budget, since the accounting for the individual fund allocations is at the discretion of the library. Nonetheless, if the various book funds were allocated on some rational basis, each fund should be totally spent at the end of the fiscal year.

What is needed is a simple method of providing the capability to predict at any time in the course of the fiscal year what the spent figure for each fund number will be at the end of that fiscal year, based on the spent figure to date and the orders that have been placed but not yet received. That is, for each item on order for each fund number it should be possible to predict how

many will be received and paid by the end of the fiscal year. What is the model?

THE SYSTEM

Before a model can be developed, the system must be defined so that an empirically derived statistical table can be built for the history of orders placed, which includes all the possible occurrences to an order. The system will apply only to firm orders, blanket orders and approval plans being excluded. Thus the scope of this model is effectively limited to those institutions where firm orders constitute a significant proportion of their monograph expenditures.

The model described may be applicable to approval plans where selection is done from agent-supplied approval forms, but since this method receives insignificant use at the University of Guelph library, it is omitted. For the present, serial and continuation orders and the problem of partial shipments are ignored. What now remains are single-item orders for which the following may occur at any given time:

(a) the ordered item is received;

(b) some status report is received for the agent, publisher, etc.;

(c) for any of a number of reasons, the order is cancelled;

(d) an item is received on the order but is incorrect, a faulty copy, wrong edition, etc.;

(e) the item is not received.

The first possibility is the statistically triv-

Tim Sauer is assistant division head, processing division, McLaughlin Library, University of Guelph, Guelph, Ontario. The author is indebted to S. S. Swaminathan of the University of Guelph Computer Services for assistance with the statistical formulations.

ial case but the whole object of the acquisition process. Assume that the remaining cases can all be reduced, for statistical purposes, to two possibilities, namely, the order is cancelled, or the order is not received. That cases (b) and (d), above, reduce to cancellation or nonreceipt follows since a status report (b) can result in three actions:

1. The order is cancelled and dropped completely or worked on some more and reordered, when it becomes a new order, reentering the system the same as any other order.

2. The status report is noted and nothing is done. If past statistics included this occurrence along with other slow items, then the present case is taken care of.

3. The status report takes the order date back to day one and therefore uses the status report date as the order date for calculations from then on.

An incorrect or faulty item, such as (d), above, can be accepted and a reorder placed for the item, whereupon the reordered item is in essence a new order and treated as such; or it can be rejected and returned. In this instance the order is handled by using the return date as the date for calculation purposes or ignoring the date and including this time delay in the statistical sample on which the calculations are made.

THE MODEL

Define:

A = Event—the order is received
B = Event—the order is outstanding
C = Event—the order is cancelled

For statistical purposes, these three events must be mutually exclusive. That is, the occurrence of one event must preclude the possible occurrence of either of the other two events. The only exception, of the library cancelling an order and the book being received subsequently, is due to an error of communication and is not a breach of the statistical requirement for mutual exclusion.

The system that is required is one that will provide the probability that a given order, which has been on order for a certain number of time units, will be received within a certain number of future time units (i.e., the end of the fiscal year).

Define:

t = time units (days, weeks, months, or whatever)

$P_n (A/t=i)$ = Probability that event A (order received) will happen by age n time units, given that the order is already at age i time units where $n>i$.

$P_n (B/t=i)$ = Probability that event B (order will not be received) will happen by age n time units, given that the order is already at age i time units where $n>i$.

$P_n (C/t=i)$ = Probability that event C (order will be cancelled) will happen by age n time units, given that the order is already at age i time units where $n>i$.

Obviously, $P_n (A/t=i) + P_n (B/t=i) + P_n (C/t=i) = 1$, since the sum of the probabilities of an order being cancelled, received, or not received for a given time period must be one.

Table 1 is a hypothetically simple history of ten orders. For convenience, call the time units "days." The probability of any one of the orders placed on the first day being received by the end of the fifth day is .8, since two orders are cancelled. The probability of an order, which has not yet been received or cancelled at the end of the third day arriving on the fourth day would be .5, since there are only four orders outstanding and two arrived the next day. The probability of an order not received or cancelled at the end of the second day being received by the end of the fourth day is 5/7, since there were seven active orders at the end of the second day and three arrived on the third day while two more arrived on the fourth day.

TABLE 1

SIMPLE HISTORY OF TEN THEORETICAL ORDERS

Time units (Days)	A (Received)	B (Outstanding)	C (Cancelled)
1	0	10	0
2	2	7	1
3	3	4	0
4	2	1	1
5	1	0	0

In general, the equation is given by (1), below:

$$P_n (A/t=i) = \frac{\sum\limits_{t=i+1}^{n} A_{i+1}}{B_i} \qquad (1)$$

where B_i = number of orders not received at day i

A_i = number of orders received at day i

Table 2 is a more complicated example based on the orders placed with one library agent in a three-month period. The choice of analyzing a sample by agent will be discussed later. Since the data were collected manually, the time period used was half-monthly where biweekly would have been better, and all remaining twelve orders were assumed to be cancelled after 9.5 months. (In fact they were not, although in an automated system they would be at some point.)

On the basis of the sample illustrated in table 2, the probability that any or all orders currently active with this agent will be received can be calculated for a given date in the future. If the end of the fiscal year were three months away (i.e., six half-months) then the probability that orders five, eight, and ten half-months old will be received before that date is as follows:

TABLE 2

BOOK AGENT'S SUPPLY HISTORY FROM A
THREE-MONTH SAMPLE

t (½ month)	A (received)	B (not received)	C (cancelled)
0	0	1735	0
1	8	1726	1
2	98	1627	1
3	430	1194	3
4	452	737	5
5	262	468	7
6	123	339	6
7	64	269	6
8	48	213	8
9	44	162	7
10	29	127	6
11	24	96	7
12	22	68	6
13	11	53	4
14	7	43	3
15	2	38	3
16	4	31	3
17	7	20	4
18	4	13	3
19	1	0	12

$$P_{11} (A/t=5) = \frac{\sum\limits_{t=6}^{11} A_{i+1}}{B_5} =$$

$$\frac{A_6+A_7+A_8+A_9+A_{10}+A_{11}}{468} =$$

$$\frac{123+64+48+44+29+24}{468} =$$

$$\frac{332}{468} = .709 \qquad (2)$$

$$P_{14} (A/t=8) = \frac{\sum\limits_{t=9}^{14} A_{i+1}}{B_8} =$$

$$\frac{44+29+24+22+11+7}{213} = \frac{137}{213} = .643 \qquad (3)$$

$$P_{16} (A/t=10) = \frac{\sum\limits_{t=11}^{16} A_{i+1}}{B_{10}} =$$

$$\frac{24+22+11+7+2+4}{127} = \frac{70}{127} = .551 \qquad (4)$$

USES

An especially significant calculation is the probability that an order placed today will be received by the end of the fiscal year. This is shown in calculation (5), below, where the fiscal year end is six half-months away.

$$P_6 (A/t=0) = \frac{\sum\limits_{t=1}^{6} A_{i+1}}{B_0} =$$

$$\frac{8+98+430+452+262+123}{1735} =$$

$$\frac{1373}{1735} = .791 \qquad (5)$$

By doing this calculation using equation (1) for every order that is still outstanding, multiplying each probability by the respective estimated cost for that item, and then totalling these figures by fund number and adding in the actual spent figure to date for that fund, a probable spent figure (*PSF*) for

the end of the fiscal year is obtained. The value of this figure is immediately obvious. It tells us the amount of money that must be committed in order to spend the budget to 100 percent by the end of the fiscal year, as shown in equation (6), below:

$$U = \text{Allocation} - (PSF + ASF) \qquad (6)$$

where U = Probable amount of money not spent (or overspent) at the end of the fiscal year (probable fiscal end balance)

PSF = Probable spent figure from adding the probable spent figure for each outstanding order in that allocation

ASF = Actual spent figure for that allocation to date

The probable fiscal end balance, if negative, implies that new ordering should stop for the time being; if positive, implies that more orders should be processed. In fact, given the results of equation (5), above, and the average cost of a book for that fund account, the approximate numbers of orders that must be placed immediately to spend that fund completely is given by equation (7):

$$N = \frac{U}{P_f\,(A/t=0)\; x\; Av} \qquad (7)$$

where N = number of orders to be placed immediately

$P_f(A/t=0)$ = Probability of an order, placed immediately, being received by the end of the fiscal year, f time units away.

Av = Average cost of a book for that fund number.

The usefulness of such figures for all fund numbers would be of great assistance to acquisition librarians who are constantly attempting to optimize the use of limited staff throughout the year.

LIMITATIONS

The above approach ignores three types of orders, as mentioned at the beginning of the argument: partial shipments, continuation orders, and serial orders.

Partial shipments must be a part of the system by virtue of the fact that it is not known that a given order will be a partial shipment until it is received. The part of the order that arrives is taken care of, and the remaining part that is to follow can either be considered as a new order or as a continuation of the old order. Experience seems the best criterion for judging, and provided that the same choice is always used in a given order type grouping, the statistics will reflect this occurrence.

Neither continuation orders nor serial orders can be accommodated by the present model. However, a similar model seems possible for both of these types of orders, based on the past received patterns for each order and reducing the total encumbered figure for each fund account by some statistically sound procedure. Possible models for these areas are being explored.

The use of the above method is based on the assumption that the history of receipt of orders in the past will be repeated in the future. Obviously, orders may be grouped by type in order to maximize this possibility. A relatively large number of orders, however, is needed in as short a time period as possible to minimize changes with time and to keep the group statistically significant. There are at least six possible factors that may be considered in the groupings:

1. fund number
2. publisher
3. agent or vendor
4. country of imprint
5. special type of order (e.g., confirmations, music, phonodisc, AV, etc.)
6. imprint date

Every library operates somewhat differently so that the following procedures outlined for the University of Guelph library may not apply at other libraries. At Guelph only one small approval plan is used, and the bulk of the ordering is by firm orders. All orders are checked for availability before ordering so that imprint date should be relatively unimportant. The University of Guelph library uses agents for various countries and is currently using two agents for North American imprints. Monitoring of the two North American agents suggests that agent is a more significant grouping than the others for the following reasons:

1. The two agents have differing average turnaround times, which vary for the same publishers. This is not surprising since the

turnaround time is the total of the transportation, communication, publisher, and the agents' efficiencies. When dealing with the same publisher, the efficiencies of all except the agent should be the same, given that both agents are in the same country and use the same method of transportation.

2. While some publishers account for a large volume, many account for too small a volume to be statistically significant.

3. Fund numbers can be used to order any type of material from anywhere in the world, and many are too small to produce a statistically significant sample.

For non–North American and British imprints, smaller volume agents can perhaps be grouped together for statistical purposes. As all audiovisual (AV) and confirmation orders are ordered direct, these may be treated either by experimentally determining similar groups or by manually assigning a probability when the order is placed (e.g., one for a reserved confirmation order, except near the end of the fiscal year).

This approach is based on an understanding of the acquisitions process. It may well happen that an implemented system would suggest the groupings be made on a different basis, which should be possible in a flexibly designed automated system.

Given the fact that more than one group of order types is likely to be needed in a given library, equation (7) would have to be generalized by summing the weighted probabilities of each statistical group as in (8) below.

$$N = \cfrac{U}{Av \; x \; \sum\limits_{j=1}^{n} W^j P^j_f \; (A/t=0)} \qquad (8)$$

where W^j = Weighting factor for each of the n statistical groups such that

$$\sum_{j=1}^{n} W^j = 1$$

and $P^j_f (A/t=0)$ = Probability of an order in statistical group j, placed immediately, being received by the end of the fiscal year, f time units away.

The model outlined is being incorporated into the new acquisition system currently in the design stage at the University of Guelph library. It will be fully integrated with the currently operational circulation system, an on-line, real-time system employing a chained file structure and running on a distributed processor minicomputer. Implementation of the acquisition system is expected for late 1978 or early 1979, to be followed by the inclusion of a new cataloging system.

This model can also be developed statistically using transitional probabilities. The mathematical formulation is far more complex looking and hence more difficult for a nonstatistician to understand. Furthermore, the transitional probability formulation would be more difficult to program and inherently more inaccurate in a computerized acquisitions system.

Since the equation derived from the transitional probability approach collapses to equation (1), the simpler development was used. Apparently, the more complex formulation must be used to check limits of confidence or tests of significance, which would give useful information on optimum sample size and time intervals to be used for various order groupings.

The model developed applies to monograph allocations where the orders placed against it are firm orders. Serial and continuation orders and blanket and most approval order plans cannot be handled by the present model. A similar statistical model seems reasonable for serial and continuation orders. The author's experience does not permit him to postulate the possibility of the development of a model for blanket and approval purchasing.

REFERENCE

1. Daniel Melcher, *Melcher on Acquisition* (Chicago: American Library Assn., 1971), p.38; Louis Round Wilson and Maurice F. Tauber, *The University Library* (2d ed.; New York: Columbia Univ. Pr., 1956), p.103–5.