DOCUMENT RESUME

ED 048 639 EA 003 349

AUTHOR	Mann, Stuart H.
TITLE	Least-Cost Decision Rules for Dynamic Information Management. Working Paper.
PUB DATE	[70]
NOTE	16p.; Paper presented at Operational Research Society of America National Meeting (38th, Detroit, Michigan, October 28-30, 1970)
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Cost Effectiveness, Information Retrieval, Information Storage, *Information Utilization, *Mathematical Models, *Technical Reports

ABSTRACT

Least cost decision rules for transferring documents from primary to secondary storage are developed from a dynamic programing model of an information system. The program is constrained to provide for a minimum acceptable level of user benefits. Knowledge of the physical size of the primary storage area and the fraction of documents returned from secondary to primary storage in each decision period is required. Transfer, handling, and circulation costs are considered. Increase in the total size of the document collection is assumed to be an uncontrolled random process. (Author)

Least-Cost Decision Rules for Dynamic Information Management

Stuart H. Mann

Division of Man-Environment Relations
The Pennsylvania State University
University Park, Pennsylvania 16802

This research has been supported partially under National Science Foundation, Office of Science Information Grant GN-759 as administered by the School of Industrial Engineering, Purdue University.

This is a working paper. Comments are invited and should be directed to the author at the above address. Please do not reproduce or quote without the author's permission.

Presented at 38th National Meeting Operations Research Society of America, Detroit, Michigan, October 28-30, 1970.

ABSTRACT

Least cost decision rules for transferring documents from primary to secondary storage are developed from a dynamic programming model of an information system. The program is constrained to provide for a minimum acceptable level of user benefits. Knowledge of the physical size of the primary storage area and the fraction of documents returned from secondary to primary storage in each decision period is required. Transfer, handling, and circulation costs are considered. Increase in the total size of the document collection is assumed to be an uncontrolled random process.

INTRODUCTION

The information explosion of the last few years has resulted in considerable research effort being directed toward the purchase and accumulation decisions faced by information centers. Two recent papers, one by W. C. Lister[1] and the other by H. M. Gurk and J. Minker[2] address this concern.

In the Lister paper, the problem of least-cost decisions for transferring information from primary storage areas to less accessible secondary storage areas is studied. He presents several models under varied assumptions. However, the return of information from secondary to primary storage is not permitted in any of Lister's models.

In contrast, Gurk and Minker investigate the size of primary storage areas when return of information from secondary to primary storage, when certain given conditions are satisfied, is allowed to occur. However, no attempt is made to identify best storage policies.

In this paper, we will incorporate the Lister and Gurk and Minker ideas and allow information to flow in both directions. In the following sections this main structure is expounded and exploited to identify optimal storage and transfer policies.

PROBLEM STATEMENT

Consider an information center with a fixed size storage area for fast access document retrieval. We will call this area the primary storage area. In the case of a library, this area might be the stacks or in the case of a computer facility the area would be the disc or drum. There is available to this information center another storage area, called secondary storage. This storage facility is less accessible than primary storage and is assumed to be unlimited in its available storage capacity.

Into the primary storage area of this information center new documents flow from an uncontrolled random process. This can be thought of as the blanket-order system for libraries wherein virtually all published books are received and processed for the collection held in primary storage. With a fixed size primary storage area and rapid increases in document input, an imbalance soon occurs unless space is made available. Space can be made available by transferring documents from primary storage to secondary storage.

If we allow documents from secondary storage to be returned to primary storage when they meet set decision criteria, then there becomes another input source for primary storage. This input compounds the problem of an already overcrowded primary storage area.

Now, given that we must not allow the collection in primary storage to drop below some fixed critical level and that a known number or expected number of documents from secondary storage are returned each decision period, the decision that must be made is how many documents do we transfer from primary to secondary storage in each decision period. It is assumed that decisions are made at the beginning of time periods of equal length, perhaps monthly.

The forces inhibiting arbitrary transfer of documents from one storage area to the other are the inherent costs involved. There are four major costs that we will consider. There are two costs involved with the circulation and handling of documents. One cost is the charge to the system for the circulation and handling of documents in the primary storage area expressed as a function of the number of documents contained therein. The other cost is the corresponding charge to the system for circulation and handling of documents in the secondary storage area. The other two costs are realized upon the transfer of documents. One charge is made for the transfer of documents from primary to secondary storage, and the other charge is for the transfer of documents from secondary to primary storage. Both of these costs are expressed as functions of the number of documents transferred.

The objective can now be stated as follows: find the number of documents to transfer to secondary storage each decision period so as to minimize the total of circulation, handling, and transfer costs in maintaining the primary and secondary collections given that the size of the primary collection must be no less than some minimum acceptable level. In the following section we will develop the mathematical model for the described system and describe the form of the optimal policy under given conditions.

MATHEMATICAL MODEL

The following necessary assumptions are made: (a) the fraction of secondary storage documents that are transferred to primary storage in each period is a fixed known value; (b) documents to be transferred from primary to secondary storage will be moved on the basis of age, the oldest moving first;[†] (c) documents moved from secondary to primary storage are considered as new documents in primary storage.

The following parameters, variables, and functions are identified for subsequent use:

P – maximum workable size of primary storage area;

β – fraction of primary collection that must be maintained to insure minimum level of user acceptability ($0 < \beta < 1$);

η – fraction of secondary collection that is transferred to primary storage in each decision period ($0 < \eta < 1$);

ξ – random variable for the number of new documents as input to the primary collection from an external source;

$\phi(\xi)$ – probability mass function for the random variable ξ, $\xi = 0,1,2,\ldots,n$;

χ – size of the primary collection at the beginning of a decision period, just before a transfer decision is made;

ω – size of the secondary collection at the beginning of a decision period, just before a transfer decision is made;

y_j – number of documents transferred from primary to secondary storage at the beginning of decision period j, $j = 1,2,\ldots,n$.

† It may be desirable to include an external provision that would maintain a document in primary storage if it had experienced considerable use, even if it were eligible for transfer to secondary storage.

The costs imposed on the system are as follows:

- p - circulation and handling charge per document in primary storage;
- s - circulation and handling charge per document in secondary storage;
- t_1 - cost per document transferred from primary to secondary storage;
- t_2 - cost per document transferred from secondary to primary storage, with all p, s, t_1 and $t_2 > 0$.

Since decisions are to be made at the beginning of each of n equal length decision periods, it is convenient to model the process as a dynamic program.[3] We will number the decision periods backward in the usual way. Let $f_j(\chi,\omega)$ be defined as the minimum total expected cost with j decision periods remaining, starting with χ documents in primary storage and ω documents in secondary storage. Define $f_0(\chi,\omega)$ to be zero for all χ and ω.

The time sequence of events is as follows: decision period begins with state of system observed as document levels in primary and secondary storage; decisions made simultaneously for transfer of documents; documents transferred; costs charged on new document levels and amounts transferred; random input into primary storage; end of decision period. For a single decision period we have

$$f_1(\chi,\omega) = \min_{y_1} \left\{ p(\chi + \eta\omega - y_1) + s(\omega + y_1 - \eta\omega) + t_1 y_1 + t_2 \eta\omega \right\} \quad (1)$$

where the decision variable y_1 must satisfy

$$\beta P \leqq \chi + \eta\omega + E(\xi) - y_1 \leqq P \tag{2}$$

and

$$y_1 \geqq 0. \tag{3}$$

Since the decision for the transfer of documents to secondary storage must take into account the maximum size of the primary storage area, the knowledge of what is to be received by primary storage during the decision period is incorporated into constraint (2). This constraint forces the transfer of documents to be large enough to enable the size of the primary collection not to exceed its upper bound at any time during the decision period. At the same time constraint (2) requires that a minimum size primary collection be maintained. Constraint (3) states that negative amounts cannot be transferred.

Now, if the objective function (1) is rewritten as

$$f_1(\chi,\omega) = \min_{y_1} \left\{ (s - p + t_1)y_1 + p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega \right\} \tag{4}$$

subject to (2) and (3), it is obvious that the optimal policy for this single decision period depends only on the coefficients p, s, and t_1. There are three cases to consider: Case (1) $s > p$; since $t_1 > 0$ and $s > p$, $s - p + t_1 > 0$ and this positive coefficient implies that y_1 should be made as small as possible, i.e., $\max[0, \chi + \eta\omega + E(\xi) - P]$. Case (2) $p > s$, $s + t_1 > p$; again this implies $s - p + t_1 > 0$ which yields the same optimal policy as case (1). Case (3) $p > s + t_1$; this implies $s - p + t_1 < 0$ which indicates that y_1 should be made as large as possible, i.e., optimal $y_1 = \chi + \eta\omega + E(\xi) - \beta P$.

For a decision process of n periods duration, we have the following objective function:

$$f_n(\chi,\omega) = \min_{y_n}\Big\{(s - p + t_1)y_n + p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega$$

$$+ \sum_{\xi} f_{n-1}(\chi + \eta\omega + \xi - y_n,\ \omega + y_n - \eta\omega)\phi(\xi)\Big\} \qquad (5)$$

subject to

$$\beta P \leqq \chi + \eta\omega + E(\xi) - y_n \leqq P$$

and $y_n \geqq 0$.

Lemma 1.

The function $f_n(\chi,\omega)$ is linear in χ and ω, for all n.

Proof.

We have assumed $f_0(\chi,\omega) = 0$ for all χ and ω.

$$f_1(\chi,\omega) = \min_{y_1 \epsilon Y_1}\Big\{(s - p + t_1)\, y_1 + p\chi + [(p+t_2)\eta + s\,(1-\eta)]\omega\Big\}$$

$$\text{where } Y_j = \Big\{y_j : \beta P \leqq \chi + \eta\omega + E(\xi) - y_j \leqq P \cap y_j \geqq 0,\ j=1,2,\dots,n\Big\}$$

Since the quantity in brackets {} is linear in y_1, the optimal y_1, $y_1^* = \max[0, \chi + \eta\omega + E(\xi) - P]$ or $\chi + \eta\omega + E(\xi) - \beta P$. In the case that $y_1^* = 0$, then $f_1(\chi,\omega) = p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega$ which can be written as $K_1 + L_1\chi + M_1\omega$ where $K_1 = 0$, $L_1 = p$, and $M_1 = [(p + t_2)\eta + s(1 - \eta)]$. In the case that $y_1^* = \chi + \eta\omega + E(\xi) - P$, $f_1(\chi,\omega) = (s - p + t_1)[\chi + \eta\omega + E(\xi) - P] + p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega$, which can be written as $K_1 + L_1\chi + M_1\omega$ where $K_1 = (s - p + t_1)[E(\xi) - P]$; $L_1 = s + t_1$, and $M_1 = (t_1 + t_2)\eta + s$. When $y_1^* = \chi + \eta\omega + E(\xi) - \beta P$, $f_1(\chi,\omega) = (s - p + t_1) \times [\chi - \eta\omega + E(\xi) - \beta P] + p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega$, which can be written as $K_1 + L_1\chi + M_1\omega$ where $K_1 = (s - p + t_1)[E(\xi) - \beta P]$, $L_1 = s + t_1$, and $M_1 = (t_1 + t_2)\eta + s$. In each case $f_1(\chi,\omega)$ is linear in χ and ω. As an induction assumption, assume $f_{n-1}(\chi,\omega) = K_{n-1} + L_{n-1}\chi + M_{n-1}\omega$ where K_{n-1}, L_{n-1}, and M_{n-1} are functions of s, p, t_1, t_2, η, β, P and $E(\xi)$ only. Since $\chi_{n-1} = \chi_n + \eta\omega + \xi - y_n$ and $\omega_{n-1} = y_n + (1 - \eta)\omega_n$,

$$f_n(\chi,\omega) = \min_{y_n \varepsilon Y_n} \Big\{ (s - p + t_1)\, y_n + p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega + \sum_{\xi} [K_{n-1} + L_{n-1}(\chi + \eta\omega + \xi - y_n) + M_{n-1}(y_n + (1 - \eta)\omega)]\phi(\xi) \Big\}$$

$$= \min_{y_n \varepsilon Y_n} \Big\{ (s - p + t_1 + M_{n-1} - L_{n-1})\, y_n$$

$$+ (p + L_{n-1})\chi + [(p + t_2)\eta + s(1 - \eta) + \eta L_{n-1}$$

$$\left. + (1 - \eta)M_{n-1}]\omega + K_{n-1} + E(\xi)L_{n-1} \right\}$$

Since the quantity inside the brackets {} is linear in y_n, y_n^* must equal one of the endpoints, i.e., $\chi + \eta\omega + E(\xi) - \beta P$ or $\max[0, \chi + \eta\omega + E(\xi) - P]$.

Thus, $f_n(\chi,\omega) =$ (1): $(s - p + t_1)[\chi + \eta\omega + E(\xi) - \beta P] + p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega + K_{n-1} + \beta PL_{n-1} + [\chi + \omega + E(\xi) - \beta P]M_{n-1}$, or (2): $p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega + K_{n-1} + [\chi + \eta\omega + E(\xi)]L_{n-1} + (1 - \eta)\omega M_{n-1}$, or (3): $(s - p + t_1)[\chi + \eta\omega + E(\xi) - P] + p\chi + [(p + t_2)\eta + s(1 - \eta)]\omega + K_{n-1} + PL_{n-1} + [\chi + \omega + E(\xi) - P]M_{n-1}$.

In each of the three cases it is observed that $f_n(\chi,\omega)$ is linear in χ and ω and can be written as $K_n + L_n\chi + M_n\omega$. In case (1); $K_n = (s - p + t_1 + M_{n-1})[E(\xi) - \beta P] + \beta PL_{n-1} + K_{n-1}$, $L_n = s + t_1 + M_{n-1}$, and $M_n = (t_1 + t_2)\eta + s + M_{n-1}$. In case (2); $K_n = E(\xi)L_{n-1} + K_{n-1}$, $L_n = p + L_{n-1}$, and $M_n = (p + t_2)\eta + s(1 - \eta) + \eta L_{n-1} + (1 - \eta)M_{n-1}$. In case (3); $K_n = (s - p + t_1 + M_{n-1})[E(\xi) - P] + PL_{n-1} + K_{n-1}$, $L_n = s + t + M_{n-1}$, and $M_n = (t_1 + t_2)\eta + s + M_{n-1}$.

Lemma 2.

(a) When $s > p$, then $s - p + t_1 + M_n - L_n > 0$, for all n.

(b) When $p > s$, $s + t_1 > p$, and $t_1 < \eta(t_1 + t_2)$, then $s - p + t_1 + M_n - L_n > 0$, for all n.

(c) When $p - s > \max[t_1, \eta(t_1 + t_2)]$, then $s - p + t_1 + M_n - L_n < 0$, for all n.

Proof.

(a) The proof will be by induction. From the objective function (4) it is seen that there are two cases to consider when $n = 1$.

Case (1): $y_1^* = 0$.

$$s - p + t_1 + M_1 - L_1 = s - p + t_1 + (p + t_2)\eta + s(1 - \eta) - p$$
$$= (2 - \eta)(s - p) + t_1 + t_2\eta > 0.$$

Case (2): $y_1^* = \chi + \eta\omega + E(\xi) - P$.

$$s - p + t_1 + M_1 - L_1 = s - p + t_1 + (t_1 + t_2)\eta + s - s + t_1$$
$$= s - p + 2t_1 + (t_1 + t_2)\eta > 0.$$

As the induction assumption, assume that $s - p + t_1 + M_{n-1} - L_{n-1} > 0$. To evaluate $s - p + t_1 + M_n - L_n$, there are three cases to consider.

Case (1): $y_n^* = \chi + \eta\omega + E(\xi) - \beta P$.

$$s - p + t_1 + M_n - L_n = s - p + t_1 + (t_1 + t_2)\eta + s + M_{n-1} - s - t_1 - M_{n-1}$$
$$= s - p + (t_1 + t_2)\eta > 0.$$

Case (2): $y_n^* = 0$.

$$s - p + t_1 + M_n - L_n = s - p + t_1 + (p + t_2)\eta + s(1 - \eta) + \eta L_{n-1}$$
$$+ (1 - \eta)M_{n-1} - p - L_{n-1}$$
$$= (s - p + t_1 + M_{n-1} - L_{n-1})(1 - \eta)$$
$$+ s - p + \eta(t_1 + t_2).$$

Thus, by the induction assumption, $s - p + t_1 + M_n - L_n > 0$.

Case (3): $y_n^* = \chi + \eta\omega + E(\xi) - P$.

Same as Case (1).

Therefore, $s - p + t_1 + M_n - L_n > 0$ for all n, when $s > p$.

(b) The proof will be by induction. From the objective function (4) it is seen that when n = 1 there are two cases to consider.

Case (1): $y_1^* = 0$.

$$s - p + t_1 + M_1 - L_1 = s - p + t_1 + (p + t_2)\eta + s(1 - \eta) - p$$
$$= (s - p + t_1) + (1 - \eta)(s - p) + t_2\eta$$
$$> (s - p + t_1) + (1 - \eta)(-t_1) + t_2\eta$$
$$= s - p + t_1 + \eta(t_1 + t_2) - t_1 > 0.$$

Case (2): $y_1^* = \chi + \eta\omega + E(\xi) - P$.

$$s - p + t_1 + M_1 - L_1 = s - p + t_1 + (t_1 + t_2)\eta + s - s + t_1$$
$$= s - p + 2t_1 + (t_1 + t_2)\eta > 0.$$

As the induction assumption, assume that $s - p + t_1 + M_{n-1} - L_{n-1} > 0$. To evaluate $s - p + t_1 + M_n - L_n$, there are three cases to consider.

Case (1): $y_n^* = \chi + \eta\omega + E(\xi) - \beta P$.

$$s - p + t_1 + M_n - L_n = s - p + t_1 + (t_1 + t_2)\eta + s + M_{n-1} - s - t_1 - M_{n-1}$$
$$= s - p + t_1 + (t_1 + t_2)\eta - t_1 > 0.$$

Case (2): $y_n^* = 0$.

$$s - p + t_1 + M_n - L_n = s - p + t_1 + (p + t_2)\eta + s(1 - \eta)$$
$$+ \eta L_{n-1} + (1 - \eta)M_{n-1} - p - L_{n-1}$$
$$= (s - p + t_1 + M_{n-1} - L_{n-1})(1 - \eta)$$
$$+ s - p + t_1 + \eta(t_1 + t_2) - t_1 > 0.$$

Case (3): $y_n^* = \chi - \eta\omega + E(\xi) - P$.

Same as Case (1).

Thus, $s - p + t_1 + M_n - L_n > 0$ for all n, when $p > s$, $s + t_1 > p$ and $t_1 < \eta(t_1 + t_2)$.

(c) The proof will be by induction. From the objective function (4) it is seen that when n = 1, $y_1^* = \chi + \eta\omega + E(\xi) - \beta P$.

$$s - p + t_1 + M_1 - L_1 = s - p + t_1 + (t_1 + t_2)\eta + s - s - t_1$$
$$= s - p + \eta(t_1 + t_2) < 0.$$

As the induction assumption assume that $s - p + t_1 + M_{n-1} - L_{n-1} < 0$. To evaluate $s - p + t_1 + M_n - L_n$, there are three cases to consider.

Case (1): $y_n^* = \chi + \eta\omega + E(\xi) - \beta P$.

$$s - p + t_1 + M_n - L_n = s - p + t_1 + (t_1 + t_2)\eta + s + M_{n-1} - s - t_1 - M_{n-1}$$
$$= s - p + \eta(t_1 + t_2) < 0.$$

Case (2): $y_n^* = 0$.

$$s - p + t_1 + M_n - L_n = s - p + t_1 + (p + t_2)\eta + s(1 - \eta) + \eta L_{n-1} + (1 - \eta)M_{n-1} - p - L_{n-1}$$
$$= (s - p + t_1 + M_{n-1} - L_{n-1})(1 - \eta) + s - p + \eta(t_1 + t_2) < 0.$$

Case (3): $y_n^* = \chi + \eta\omega + E(\xi) - P$.

Same as in Case (1).

Therefore, $s - p + t_1 + M_n - L_n < 0$ for all n, when $p - s > \max[t_1, \eta(t_1 + t_2)]$.

<u>Theorem</u>

The optimal transfer policy in each period of an n-period process takes the following form (a) when either $s > p$ or $p > s$, $s + t_1 > p$ and $t_1 < \eta(t_1 + t_2)$:

Then if $\begin{cases} \chi + \eta\omega + E(\xi) > P, \text{ transfer } \chi + \eta\omega + E(\xi) - P \text{ to secondary storage} \\ \chi + \eta\omega + E(\xi) \leq P, \text{ do nothing.} \end{cases}$

(b) When $p - s > \max[t_1, \eta(t_1 + t_2)]$, transfer $\chi + \eta\omega + E(\xi) - \beta P$ to secondary storage.

Proof

(a) From Lemma 1 it is seen that $f_n(\chi,\omega)$ can be written as

$$f_n(\chi,\omega) = \min_{y_n \varepsilon Y_n} \Big\{ (s - p + t_1 + M_{n-1} - L_{n-1}) y_n + (p + L_{n-1})\chi + [(p + t_2 + L_{n-1})\eta + (s + M_{n-1})(1 - \eta)]\omega + K_{n-1} + E(\xi)L_{n-1} \Big\}$$

Since the quantity within the brackets {} is linear in y_n, the solution must occur at an endpoint. The endpoint is dependent solely on whether the coefficient of y_n is positive or negative. By Lemma 2, parts (a) and (b), the coefficient is always positive under either of the given conditions. Thus, to minimize the quantity within the brackets {}, y_n should be made as small as possible. Therefore, $y_n^* = \max[0, \chi + \eta\omega + E(\xi) - P]$, and the optimal policy results.

(b) The reasoning is identical to part (a) with the exception that now the coefficient of y_n is negative under the given conditions as shown in Lemma 2, part (c). Thus, y_n should be made as large as possible.

Therefore, $y^*_\eta = \chi + \eta\omega + E(\xi) - \beta P$.

CONCLUSION

Sufficient conditions for simple operating rules are given by the theorem. These conditions are dependent on the cost parameters of the system and the fraction of documents returning to primary storage from secondary storage in each decision period. In addition to these parameters, implementation would require knowledge of the expected value of the number of new arrivals to the system in each period of the process.

REFERENCES

1. Lister, Winston C. "Least Cost Decision Rules for the Selection of Library Materials for Compact Storage." Unpublished Ph. D. dissertation, School of Industrial Engineering, Purdue University, Lafayette, Indiana. January, 1967.

2. Gurk, Herbert M. and Jack Minker. "Storage Requirements for Information Handling Centers." Technical Report 68-61, University of Maryland Computer Science Center, College Park, Maryland. February, 1968.

3. Bellman, Richard E. and Stuart E. Dreyfus. Applied Dynamic Programming Princeton University Press, Princeton, New Jersey, 1962.