STUDY OF A PARTICULAR MULTIPLE LEVEL INVENTORY MODEL IN CASH MANAGEMENT

Michel Cramer

Jean-Pierre Ponssard

April 1974

Study of a Particular Multiple Level Inventory Model in Cash Management*

Michel Cramer** and Jean-Pierre Ponssard ***

1. Description of the Model

The model presented here summarizes the principal characteristics of a real problem encountered during a consultancy experience. It is concerned with the cash management of a banking network comprising:

a) a central bank which can deposit or withdraw from the cash inventory at a bank exterior to the system,

b) peripheral (i.e. branch) offices which can deposit or withdraw from the cash inventory at the central bank, and

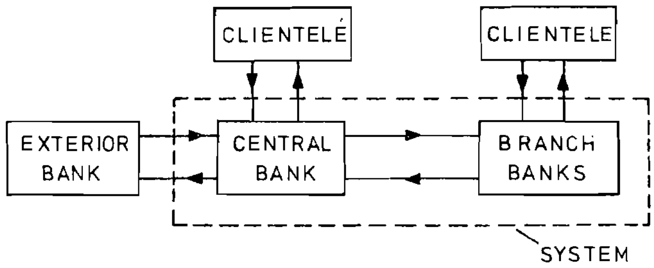
c) various client groups which deposit or withdraw from the cash inventory at the central bank and at each branch.

---

*Translated from the original French, "Etude d'un modèle particulier de gestion de stocks avec niveaux multiples." This paper was presented at the AFCET Congress, Paris, France, 15-17 May, 1974, and published in the proceedings of the Congress, Aide à la Décision, Tome 2.

**Centre d'Enseignement Supérieur du Management Public, 94112, Arcueil, France.

***On leave from Centre d'Enseignement Supérieur du Management Public, and Groupe de Recherche en Gestion des Organisations Ecole Polytechnique, Paris, France; research scholar at the International Institute for Applied Systems Analysis, Laxenburg, Austria.



For reasons of economy and security, the general objective is to minimize the amount of the cash inventory held in the system, and secondarily, to control the frequency and the total number of transactions between banks.

The study treats the following questions:

A. Coordinate the schedules of cash transactions between the banks (Section 2).

B. Transmit between the banks information about the transactions (Section 3).

C. Predict the cash transactions with the various clients (Section 4).

We will present Problems A and B above independently and under the following two aspects:

Practical--the practical result obtained during the study in relation to the problem posed.

Theoretical--the abstract models inspired by a generalization of the problem.

2. The Transaction Schedules

2.1 Practical Aspect

In the banking network studied, we first noticed that the principal transactions, i.e.

a) hourly schedule of exchanges with the small clients at the teller windows,

b) special transactions with the big clients, and

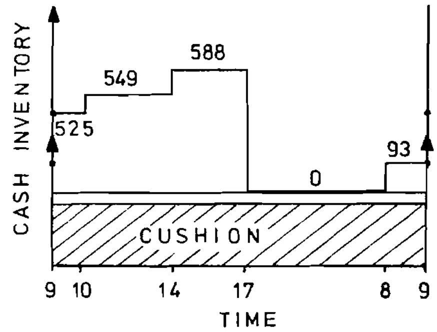
c) transfers from branch to branch,

are regulated by a constant daily schedule. The cash report of each office is thus formulated as a list of fixed hours and variable sums.

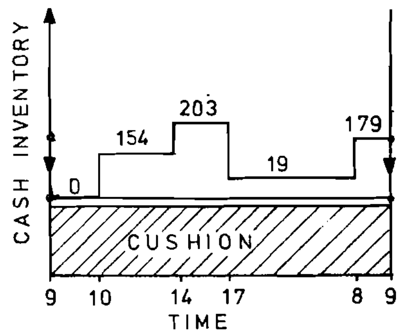
For example, the central bank presents the following cash report:

Hour of Transaction		10:00	14:00	17:00	8:00
Start of Month	Average total	24	39	-588	93
	Standard deviation	23	16	385	64
End of Month	Average total	154	49	-184	160
	Standard deviation	117	21	123	143

A transaction with the exterior bank at 9:00 permits the central bank to bring its inventory to a level equal to the average need of the day, augmented by a "cushion" to take account of the uncertainty of its total. The level of the cash inventory thus has the following average "profile":



Average profile of the cash inventory at the beginning of the mont! (withdrawal at 9:00)

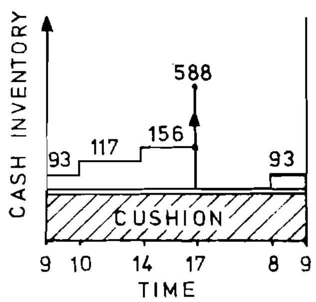


Average profile of the cash inventory at the end of the month (deposit at 9:00)

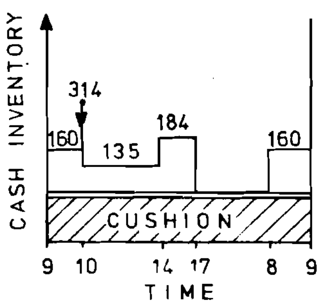
In fact, certain hours of transactions may be seen as decision variables in the sense that one can alter them within broad intervals. This is particularly the case in most of the transactions between the branches.

The practical tool proposed to those in charge of the cash inventory of the banking network to judge the appeal of a schedule change consists of comparing the average profile of the cash inventory in the two cases, and in particular the surface produced by this profile, which represents the average cash inventory ("cushion" excluded).

Application of this method to the above example is especially simple. It discloses the optimal schedule of transactions between the central bank and the exterior bank (17:00 at the beginning of the month for a withdrawal and 10:00 at the end of the month for a deposit):



Beginning of the month
(withdrawal at 17:00)



End of the month
(deposit at 10:00)

2.2 Theoretical Scheduling of the Cash Transactions Application of Dynamic Programming[1]

This study of the schedule appears as a particular case of the general problem of cash transaction scheduling formulated here from a theoretical viewpoint.

During the interval $[0,T]$, N movements of quantities $\omega = (\omega_i;\ i = 1,\dots,N)$ change the cash inventory, the initial level of which is S. The problem consists of choosing the hour Z_i of each transaction in the interval $[a_i, b_i]$ in order to minimize the average level of the cash inventory while keeping it positive at all times. It is convenient to define $(t_k;\ k = 1,\dots,2N + 1)$ as

$$t_k \leq t_{k+1} \quad , \quad t_{\underline{k}_i} = a_i \quad , \quad t_{\overline{k}_i} = b_i \quad , \quad t_{2N+1} = T$$

and

$$(T_k;\ \ k = 1,\dots,2N) \ : \ T_k = t_{k+1} - t_k \ .$$

First one notes that all solutions for which certain transactions occur at an instant Z, such that

$$t_k < Z < t_{k+1} \quad ,$$

can be improved by taking instead $Z = t_{k+1}$ if the sum of these transactions is positive, or $Z = t_k$ if it is negative. In the two cases the new solution does not increase the cash inventory and still satisfies the constraints.

---

[1]This section was developed in collaboration with Carlos Winkler.

Let x_k^i be defined as follows. If

$$t_k < Z_i \quad ,$$

then

$$x_k^i = 0 \quad ,$$

and if

$$t_k \geq Z_i \quad ,$$

then

$$x_k^i = 1 \quad .$$

One can thus solve the problem by a dynamic program:

- the state at stage $k (k = 1,\dots,2N)$ is the vector

$$x_k = (x_k^i ; \quad i = 1,\dots,N) \quad ,$$

- the economic function F_k measures the cash inventory in the interval $[t_k, T]$:

$$F_k(x_k) = \sum_{e=k}^{2N} \sum_{i=1}^{N} \omega_i x_e^i T_e \quad .$$

A_1

Stage 2N (start-up):

$$x_{2N} = (1,\dots,1)$$

$$F_{2N}(x_{2N}) = \sum_{i=1}^{N} \omega_i T_{2N} + S.T$$

Stage k (Recurrence):

$$F_k(x_k) = T_k \sum_{i=1}^{N} \omega_i x_k^i + \min_{(x_{k+1}^i \geq x_k^i)} F_{k+1}(x_{k+1})$$

for x_k such that

$$\sum_{i=1}^{N} \omega_i x_k^i + S \geq 0 \quad \text{(positive cash inventory)}$$

$$\left.\begin{array}{l} x_k^i = 1 \text{ if } \overline{k}_i \leq k \\ x_k^i = 0 \text{ if } k < \underline{k}_i \end{array}\right\} \quad \text{(transaction schedules)}$$

To take into account the case of an infinite horizon with a sequence of transactions repeating itself during the intervals [0,T], [T,2T]..., one must take

$$\sum_{i=1}^{N} \omega_i = 0$$

and consider the evolution of the minimal level s_k of the initial cash inventory which permits us to obtain a positive inventory of t_k at T:

$$s_{k+1} = s_k + \max\left[0, - \sum_{i=1}^{N} \omega_i x_k^i\right] .$$

One can thus adapt Program A_1 to deal with this case:

A_2

Stage 2N:

$$x_{2N} = (1,\dots,1)$$
$$F_{2N}(x_{2N}) = 0$$

Stage k:

$$F_k(x_k) = T.\max\,(0, - \sum_{i=1}^{N} \omega_i x_k^i) + T_k \sum_{i=1}^{N} \omega_i x_k^i + \min_{x_{k+1}^i \geq x_k^i} F_{k+1}(x_{k+1})$$

for x_k such that

$$x_k^i = 1 \text{ if } \bar{k}_i \leq k$$

$$x_k^i = 0 \text{ if } k < \underline{k}_i .$$

Remarks

a) We note that it is advantageous to calculate $F_k(x_k)$ in the decreasing lexicographic order of the possible binary vectors for x_k, so that the computation of min $F_{k+1}(x_{k+1})$ is limited to a single comparison with the preceding value.

b) Programs A_1 and A_2 contribute a noticeable economy in computation in comparison with exhaustive enumeration. Indeed, let n_k and m_i be the respectiv numbers of indices i and k such that

$$a_i \leq t_k \leq b_i \quad ;$$

the number of evaluations at stage k is 2^{n_k}, or a total of

$$\sum_{k=1}^{2N} 2^{n_k}$$

which increases linearly with N, while the number of possible solutions

$$\prod_{i=1}^{N} m_i$$

follows an exponential tendancy.

Application of the linear program. If, within the framework of the above model, one allows a transaction to be spread among several dates (which is logical if the constant share of the costs of the transactions is negligible), the problem is formulated as a linear program in the following manner (minimum flow network with sources and sinks):

$$A_3: \quad \text{Min} \sum_{k=1}^{2N} \sum_{i=1}^{N} y_k^i \, (T - t_k) + ST$$

$$\text{with } y_k^i = 0 \quad \text{for } t_k \notin [t_{\underline{k}_i}, t_{\overline{k}_i}] \qquad \text{(Schedules)}$$

$$\sum_{k=1}^{2N} y_k^i = \omega_i \qquad \text{(Spread of transaction)}$$

$$\sum_{k=1}^{l} y_k^i + S \geq 0 \qquad \text{(Positive cash inventory)}$$

Remark

This program can be generalized to take into account:

a) the limits for the hourly transactions and the cash inventory,

b) the costs per quantity of transaction and the holding costs varying according to date, and

c) multiple levels of cash inventory.

3. Communication between the Banks

3.1 The Practical Result of the Study

In the case of the central bank described in section 2.1, the study of the average profile of the cash in hand allows us to advocate a priori 17:00 for the moment of withdrawal to the external bank at the start of the month. But, the constraints of the schedule of this activity lead us in fact to adopt 14:00 (instead of 9:00 initially).

However, at the start of the month the most important and the most uncertain transaction comes at 17:00. It is concerned in fact with the messages from the branch banks communicating their needs for the next day. The corresponding cash report is thus prepared, counted by 18:00 at the latest (schedule constraint), then sent out the next morning.

Incidentally, in this period of the start of the month, the branch offices experience large irregular transactions (local clientele) from 14:00 to 18:00 with the result that the relative precision of their estimated need depends largely upon the time.

During this study, three options were contemplated for the communication system between the branch offices and the central office:

a) request of the branch banks at 18:00,
b) definitive request at 14:00, and
c) provisional request at 14:00 and final modified request at 18:00.

A series of tests for sample days and offices showed that option (c) was clearly preferable. This practical result upholds the theoretical demonstration described below.

3.2 Optimal Communication between the Banks

Let us consider the following model. At each hour t of the day ($0 \leq t \leq 1$) each branch office can give an estimator X_t of the quantity of cash necessary for the following day. To express the fact that the uncertainty progressively lessens with time, let us assume that

$$Y_{t_1}^{t_2} = X_{t_2} - X_{t_1}$$

is a normal random variable with an average of 0 and a variance of $t_2 - t_1$ (>0).

Let us further assume that $Y_{t_1}^{t_2}$ and $Y_{t_3}^{t_4}$ are independent random variables if the intervals t_1t_2 and t_3t_4 do not overlap.

We further define:

α and β are the quality of service required by a branch office and by the central office respectively;

T is the hour of the transaction between the central office and the external office which supplies it;

$\tau_1 < T$ and $\tau_2 > T$; and

n is the number of branch offices (assumed identical).

Within this framework, the various communication schedules between the branch offices and the central office are evaluated in terms of total security cash in the following table:

Communication of the branches at time:	τ_1	τ_1 and τ_2	τ_2
Total security cash in the network	$n\alpha\sqrt{1-\tau_1}$	$n\alpha\sqrt{1-\tau_2}$ $+ \beta\sqrt{n(\tau_2-\tau_1)}$	$n\alpha\sqrt{1-\tau_2}$ $+ \beta\sqrt{n\tau_2}$

The coordination of the quality of service of the branches implies

$$\alpha = \beta \quad .$$

Under these conditions solution 2 is superior to the other two solutions whatever the τ_1 and τ_2. Moreover, the optimum is obtained for $\tau_2 = 1$ and $\tau_1 = T$.

4. Predicting the Cash Need of the Central Bank

4.1 Theoretical Model

a) The composite model predicts each transaction; the global model predicts directly their aggregation.

b) Explanatory variables:

i) seasonal aspects (global habits of the clients)

ii) paydays of the big companies of the region.

c) Through use of binary variables, a linear model (multiple regression) of covariance analysis allows global estimation of the seasonal effects and other explanatory variables. As the true value of the latter are unobservable but noticeably constant, the binary variables will be substituted for them in the guise of instrumental variables.

Setting up an information system which would permit the observation of the true value of these variables will increase the explicative power of the model.

The problem of co-linearity between certain seasonal variables and those of paydays occurring almost always at the same time of the month makes imprecise the estimation of the marginal contribution of the variables in question. But this does not alter the model's explicative power so long as the relation between the two variables remains the same.

4.2 Results

By a priori analysis, by discussion with the practitioners, and by systematic examination of the residuals from the first versions of the model, the following seasonal aspects were discerned:

a) position of the day in the week and of the week in the month (WEEK model), or of the day in the month (MONTH model), and

b) effect of the end of the trimester; effect of the last day of the month; effect of the start of the summer holiday period.

Adjustment over a two-year period for the two models (WEEK and MONTH) is equivalently good (83% of explained variance). Nevertheless, the best test of such a descriptive model is its predictive power. Over the three-month period reserved for a test run, the WEEK model obtained much better results--as good for the standard deviation and the average of the absolute value of the residuals as for the number and the frequence distribution of the big residuals.

4.3 Composite Model or Global Model?

Only knowledge of the global need (aggregation of the transactions up to the moment the minimal cash inventory is achieved) is important. Since the explicative factors are the same for all the transactions, the question arises whether we should use a composite model or a global model.

For a transaction i[i = 1,n] the value achieved a posteriori will be

$$\underset{\sim}{y}_i = \hat{\underset{\sim}{y}}_i + \hat{\underset{\sim}{\varepsilon}}_i \quad ,$$

where

$\underset{\sim}{y}_i$ is the vector of ex post observations,

$\hat{\underset{\sim}{y}}_i$ is the vector of predicted values from the model, and

$\hat{\underset{\sim}{\varepsilon}}_i$ is the residuals vector: ex post observation/prediction difference.

This difference can come from a poor specification of the model, from observation errors, from random variations. Its standard deviation is a measure of the risk incurred in making the prediction $\hat{\underset{\sim}{y}}_i$.

For the global need

$$\underset{\sim}{y} = \sum_{i=1}^{n} \underset{\sim}{y}_i = \sum_{i=1}^{n} \hat{\underset{\sim}{y}}_i + \sum_{i=1}^{n} \hat{\varepsilon}_i$$

the direct prediction gives

$$\underset{\sim}{y} = \hat{y} + \hat{\varepsilon} \quad .$$

As the model is linear and the factors are the same, if we let

$$\hat{y}_i = X\, b_i \quad ,$$

where

$$b_i = (X'\, X)^{-1}\, X'\, y_i \equiv M\, y_i \quad ,$$

we obtain

$$\sum_{i=1}^{n} \hat{y}_i = X \left[\sum_{i=1}^{n} b_i \right] = X\, M \left[\sum_{i=1}^{n} y_i \right] = X \left[M\, y \right] = \hat{y} \quad ;$$

and thus also

$$\hat{\varepsilon} = \sum_{i=1}^{n} \hat{\varepsilon}_i \quad .$$

The prediction error of the global model is the sum of the errors of prediction over each of the transactions.

The variance of $\hat{\varepsilon}$ can thus be $\gtreqless$ to the sum of the variance of each of the $\hat{\varepsilon}_i$ according to their correlation. This variance is a measure of the explicative power of the model.

The standard deviation (measure of risk) of $\hat{\varepsilon}$ will be in all cases less than the sum of the standard deviations for each of the transactions. Indeed,

$$\sigma^2(\hat{\varepsilon}) = \sum_{i=1}^{n} \sigma^2(\hat{\varepsilon}_i) + 2 \sum_{i=1}^{n} \sum_{j>i} \rho_{ij}\, \sigma(\hat{\varepsilon}_i)\, \sigma(\hat{\varepsilon}_j)$$

$$\leq \left[\sum_{i=1}^{n} \sigma(\hat{\varepsilon}_i) \right]^2 \quad ,$$

since ρ_{ij} = the correlation between ξ_i and ξ_σ is lower in absolute value to 1.

The computations to make are n times more numerous for the composite model.