Dynamic Programming Under Uncertainty with a Quadratic Criterion Function

Herbert A. Simon

Econometrica, Vol. 24, No. 1 (Jan., 1956), 74-81.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28195601%2924%3A1%3C74%3ADPUUWA%3E2.0.CO%3B2-5

Econometrica is currently published by The Econometric Society.

DYNAMIC PROGRAMMING UNDER UNCERTAINTY WITH A QUADRATIC CRITERION FUNCTION

By Herbert A. Simon[1]

In the past five years, the dynamic programming problem has been formulated with considerable generality, and a great deal is now known about the conditions under which solutions exist and are unique. ([2], ch. 2; [3]). Certain procedures have been developed which "in principle" provide solutions. But in all save the simplest cases, the computational difficulties in actually obtaining numerical solutions are severe, and only a small number of special cases has been brought into the range of convenient, or even feasible, computation.

Particular interest attaches to the case of uncertainty—in which the payoff depends both on the strategy selected and on the joint probability distribution of environmental variables. When the criterion function is the sum of terms that are of the same type for all time periods, and when, in addition, the probability distribution is stationary and the product of independent distributions for each time period, the problem may be solvable. Under these conditions, we obtain integral equations of a familiar type (see, for example: [1], §4, 5; [3] §4.2–4.4). When the assumptions of stationarity and independence are relaxed, however, matters become far more complex, especially from a computational standpoint.

1. Outline of results

There appears to be a considerable range of practical dynamic programming problems in which the criterion function can reasonably be approximated by a sum of quadratic and linear terms. (See [5] and [7].) When this is so, the determination of the optimal course of action becomes, even under the assumption of uncertainty, extremely simple from a mathematical standpoint. Not only are the computations easily managed, but the exact solution can be exhibited in explicit algebraic form—something quite uncharacteristic of programming problems. The derivation of the solution for the case of certainty has been published in [7].

In the present paper I shall show that, when the criterion function is quadratic, the planning problem for the case of uncertainty can be reduced to the problem for the case of certainty simply by replacing, in the computation of the optimal first-period action, the "certain" future values of variables by their unconditional expectations. *In this sense*, the unconditional expected values of these variables may be regarded as a set of sufficient statistics for the entire joint probability distribution, or alternatively, as a set of "certainty equivalents."

To make clear exactly what this assertion means, let us consider the following

[1] Research undertaken for the project *Planning and Control of Industrial Operations*, under contract with the Office of Naval Research. I am indebted to a number of persons for helpful suggestions, and should like especially to mention A. Charnes (particularly with respect to the crucial step leading from (10) to (12) in the proof), C. C. Holt, F. Modigliani, S. Reiter, and H. Theil.

three kinds of information a decision-maker might possess about the future values of certain variables that are relevant to his decision:

(a) He might know these future values with certainty;

(b) He might know their unconditional expected values;

(c) He might know the joint probability distribution of the variables over the whole sequence of future time periods.

The initial task of the decision-maker is to determine his course of action *for the first time period.* At the end of that period, and on the basis of the new information then available to him, he chooses a course of action for the second period, and so on.

Certainty-equivalent Method. If the decision-maker knows (c), he can compute the expected values of (b), and (if he wishes to do so) he can behave *as if* these expected values were the (unknown) certain values of (a). If he does so, he has transformed his problem into one of dynamic programming under certainty. Having solved the latter problem, he can take the action for the first period that is indicated as optimal by this procedure. At the end of the first period, he will have new initial conditions and a new joint probability distribution, (c), from which he can obtain new expected values (b), which he can again employ in place of (a) in choosing a course of action for the second period; and so on. *It is important to observe that the decision-maker, in order to apply this procedure, does not in fact need to know* (c), *but only* (b).

General Programming Method. Alternatively, the decision-maker, again knowing (c) but not (a), can determine what action is optimal for the first period in the light of his complete knowledge of the joint probability distribution; can carry out this action; and can then replan at the beginning of the second period by the same method. With this procedure, the decision-maker does in fact need to know (c).

By definition, when the decision-maker's information consists of (c), no planning procedure can yield a higher (or "lower" if he is minimizing) expected value of the criterion function than the general programming method. What we shall show here is that the certainty-equivalent method will lead to exactly the same prescription of action as the general programming method; and hence that the former, which requires knowledge only of (b), represents a solution to the problem of dynamic programming under uncertainty with a quadratic criterion function.[2]

It is worth noting that we shall not require assumptions of stationarity or independence for the joint probability distribution of the stochastic variables. On the other hand, we shall have to pay for this generality by making strong assumptions about the form of the criterion function—specifically, that it is a quadratic form in the variables and their derivatives.[3] In another publication,

[2] This result was obtained by H. Theil [9] for the static case of planning under uncertainty with a quadratic criterion function. Unfortunately, his very elegant and simple proof does not extend in any obvious way to the dynamic case, and I have been forced to use a somewhat more complicated line of attack.

[3] This is a sufficient, not a necessary, condition for the result to hold. Generalizations will be discussed at the end of the paper.

[8], it has been shown that there is a fundamental connection between the quadratic character of the criterion function and the linearity of the resulting optimal decision rule. Hence, the assumption made explicitly here is no more restrictive than the assumption made implicitly in servomechanism design when the range of admissible alternatives is limited to linear systems.

2. PROBLEM IN INVENTORY AND PRODUCTION CONTROL

No attempt will be made here to achieve the greatest possible generality, even within the limits just set. The proof will be carried through for a system representing an inventory and production control problem of fairly wide application. The nature of the possible generalizations will be indicated briefly at the end of the paper.

The specific problem is to determine at the beginning of each month the quantity, $P(1)$, of a commodity to be produced by a factory during that month. An estimate of future sales, $S(t)$, $t = 1, \cdots, N$, is available in the form of a joint probability distribution of sales for N succeeding months. (Alternatively, we may specify: "Forecasts of sales are subject to random errors, whose joint probability distribution is known.") The initial inventory and production levels, $I(0)$ and $P(0)$, are fixed and given; but the terminal inventory and production levels, $I(N)$ and $P(N + 1)$, are not fixed.

Costs are associated with each period as a function of: (1) the inventory held during the period (costs of holding inventory and of run-outs resulting from inadequate inventory); (2) the level of production (costs associated with the static production function); and (3) the change in level of production (costs of hiring, training, termination, etc.). Each of these three components of the cost function is assumed to be quadratic. The expected value of total cost over the N periods is to be minimized, by determining the optimal policy or strategy.

The proof we shall use to show that the certainty-equivalent method is optimal is straightforward, if somewhat complicated in application. We shall first formulate the problem of determining an optimal policy by the general programming method (equation (8)), and then derive certain conditions that such a policy must satisfy. In particular, we shall show that the optimum production for the first period can be a function only of the expected values of sales, and is independent of other parameters of the joint probability distribution of sales (equation (12)). It follows that the certainty-equivalent method prescribes the same first-period production as the general programming method. Moreover, the computations required to apply the former method are extremely simple (see [7]).

3. PROOF OF THE MAIN RESULT

We associate with each time period the costs incurred "in that period," i.e.,

$$\gamma_t = C_I[I(t) - I_c]^2 + C_P[P(t) - P_c]^2 + C_{P\Delta}[P(t) - P(t-1)]^2 \qquad (t = 1, \cdots, N), \tag{1}$$

where C_I, C_P, $C_{P\Delta}$, I_c, and P_c are known constants; and we postulate that the

total cost we wish to minimize is the sum of these period costs over N periods, with the terminal conditions unspecified:

$$
\begin{aligned}
C = \sum_{t=1}^{N} \gamma_t &= C_I \sum_{t=1}^{N} [I(t) - I_c]^2 + C_P \sum_{t=1}^{N} [P(t) - P_c]^2 \\
&\quad + C_{PA} \sum_{t=1}^{N} [P(t) - P(t-1)]^2 \\
&= C_I \sum_{t=1}^{N} \left\{ \sum_{j=1}^{t} P(j) - \sum_{j=1}^{t} S(j) + I(0) - I_c \right\}^2 + C_P \sum_{t=1}^{N} [P(t) - P_c]^2 \\
&\quad + C_{PA} \sum_{t=1}^{N} [P(t) - P(t-1)]^2
\end{aligned} \tag{2}
$$

where the last substitution follows from the definition of the I's:

$$
I(t) = I(t-1) + P(t) - S(t) \qquad (t = 1, \cdots, N). \tag{3}
$$

Ignoring the terminal conditions can be justified on the basis of the theorem of Dvoretsky, *et al.*, ([3] §4.1), that a policy that is optimal for N periods incurs a total cost that is within ϵ, however small, of a policy optimal for an infinite span, provided only that N is sufficiently large, and that the cost function satisfies certain boundedness and convergence conditions. Convergence is assured in the present case through the terms in $I(t)$—because of inventory costs, production will not be undertaken in anticipation of sales indefinitely far in the future—even in the absence of a discounting factor for future costs.

We define a *policy* (or strategy) as a set of functions, $\theta_i[S(1), \cdots, S(i-1)]$ $i = 1, \cdots, N$, such that for each period we set $P(i) = \theta_i$. This means that we do not select a specific sequence of P's, but a sequence of *functions*, so that only $P(1)$ is definitely determined by the policy, and the P's for subsequent time periods may depend upon previous sales. Thus, the planner fixes only his production for the first period; on the basis of the policy, production for the second period is a function of sales for the first period, the specific amount to be determined when that information becomes available.

Actually the assumption is more general than this, since production can be made to depend on previous productions, on starting inventories, and on previous sales in any desired way, and then the latter two variables solved out iteratively (starting with $P(1)$) to obtain the θ function in the given form. The initial inventory and production levels may also appear in the θ functions, but because they are given constants they do not need to be exhibited explicitly in the argument that follows.

We define also the *optimal policy*:

$$
P^*(i) = \phi_i[S(1), \cdots, S(i-1)] = \phi(i) \quad (i = 1, \cdots, N), \tag{4}
$$

as the set of θ's that will minimize total expected cost over the joint probability

distribution of the S's. (For convenience of notation, we also define $P^*(0) = \phi(0) = P(0)$.) For *any* set $\{\theta\} \neq \{\phi\}$, we may write without loss of generality:

$$P(i) = \theta_i = \phi_i[S(1), \cdots, S(i-1)] + \epsilon\eta_i[S(1), \cdots, S(i-1)] \quad (i = 1, \cdots, N), \tag{5}$$

where ϵ is an arbitrary positive constant independent of i. Equation (5) simply defines a new set of functions, the η's. We use the operator $\mathcal{E}$ to designate the expected values of variables as viewed at the beginning of the first time interval. We wish to consider the minimum of:

$$\mathcal{E}C = \int \cdots \int Cf_N[S(N) \mid S(N-1), \cdots, S(1)], \cdots, f_1[S(1)]\, dS(N) \cdots dS(1). \tag{6}$$

Here $f_i\,[S(i) \mid S(i-1), \cdots, S(1)]$ represents the conditional probability density function for $S(i)$ for given values of the previous S's. The cost function, C, under the integral signs is a function of the policy functions, the values of which in turn depend on the S's. We assume that term-by-term integration is possible, and hence that $\mathcal{E}C$ can be expressed as a sum of the expected values of the $3N$ terms that comprise C. For a given policy, $\{\theta\}$, we have, from (5) and (3),

$$\begin{aligned}
\mathcal{E}C = C_I \sum_{t=1}^{N} \mathcal{E}\Bigg\{ & \left[\sum_{j=1}^{t} \phi(j)\right]^2 + \left[\epsilon \sum_{j=1}^{t} \eta(j)\right]^2 + 2\left[\sum_{j=1}^{t} \phi(j)\right]\left[\epsilon \sum_{j=1}^{t} \eta(j)\right] \\
& + \left[\sum_{j=1}^{t} S(j)\right]^2 + I^2(0) + I_c^2 - 2\left[\sum_{j=1}^{t} \phi(j)\right]\left[\sum_{j=1}^{t} S(j)\right] \\
& - 2\epsilon\left[\sum_{j=1}^{t} \eta(j)\right]\left[\sum_{j=1}^{t} S(j)\right] + 2(I(0) - I_c) \sum_{j=1}^{t} \phi(j) \\
& + 2(I(0) - I_c)\, \epsilon \sum_{j=1}^{t} \eta(j) - 2(I(0) - I_c)\left[\sum_{j=1}^{t} S(j)\right] - 2I(0)I_c \Bigg\} \\
& + C_P \sum_{t=1}^{N} \{\mathcal{E}\phi^2(t) + 2\,\epsilon\,\mathcal{E}[\phi(t)\eta(t)] + \epsilon^2\mathcal{E}\eta^2(t) + P_c^2\} \\
& - 2C_P \sum_{t=1}^{N} P_c\,\mathcal{E}\phi(t) - 2C_P\epsilon P_c \sum_{t=1}^{N} \mathcal{E}\eta(t) + C_{PA} \sum_{t=1}^{N} \mathcal{E}[\phi(t) - \phi(t-1)]^2 \\
& + C_{PA}\,\epsilon^2 \sum_{t=1}^{N} \mathcal{E}[\eta(t) - \eta(t-1)]^2 \\
& + 2\epsilon\, C_{PA} \sum_{t=1}^{N} \mathcal{E}\{[\phi(t) - \phi(t-1)][\eta(t) - \eta(t-1)]\}.
\end{aligned} \tag{7}$$

Now, if C^* is $C\{\phi_i\}$, then $\mathcal{E}C = \mathcal{E}C^* +$ the terms in (7) involving $\{\eta_i\}$, i.e.:

$$\begin{aligned}\mathcal{E}C = \mathcal{E}C^* &+ C_I \sum_{t=1}^{N} \mathcal{E}\left\{\epsilon^2\left[\sum_{j=1}^{t}\eta(j)\right]^2 + 2\epsilon\left[\sum_{j=1}^{t}\phi(j)\right]\left[\sum_{j=1}^{t}\eta(j)\right]\right. \\ &\left. - 2\epsilon\left[\sum_{j=1}^{t}\eta(j)\right]\left[\sum_{j=1}^{t}S(j)\right] + 2(I(0) - I_c)\,\epsilon\sum_{j=1}^{t}\eta(j)\right\} \\ &+ C_P \sum_{t=1}^{N}\{2\,\epsilon\,\mathcal{E}[\phi(t)\eta(t)] + \epsilon^2\mathcal{E}[\eta^2(t)]\} - 2C_P\,\epsilon\,P_c\sum_{t=1}^{N}\mathcal{E}\eta(t) \\ &+ C_{PA}\,\epsilon^2\sum_{t=1}^{N}\mathcal{E}[\eta(t) - \eta(t-1)]^2 \\ &+ 2\,\epsilon\,C_{PA}\sum_{t=1}^{N}\mathcal{E}\{[\phi(t) - \phi(t-1)][\eta(t) - \eta(t-1)]\}.\end{aligned} \tag{8}$$

Since $\mathcal{E}C^*$ is the minimum of $\mathcal{E}C$ for all $\{\theta\}$, the sum of the remaining terms on the right-hand side of (8) must be positive for all positive ϵ and all admissible sets of functions $\{\eta\}$. "Admissible" means here only that η_i is a function of, at most, $S(1), \cdots, S(i-1)$, and that $\eta_0 = 0$. (See remark following equation (4).) As $\epsilon \to 0$, the terms that are linear in ϵ will dominate the terms in ϵ^2. If the sum of the linear terms in ϵ is positive for given $\{\eta\}$, it will be negative for $\{-\eta\}$, and the latter functions will be admissible if the former are. Hence a necessary condition that $\mathcal{E}C^*$ be a minimum is that the sum of the terms linear in ϵ be zero. That is:

$$\begin{aligned}\Psi = C_I \sum_{t=1}^{N}\mathcal{E}&\left\{\left[\sum_{j=1}^{t}\phi(j)\right]\left[\sum_{j=1}^{t}\eta(j)\right] - \left[\sum_{j=1}^{t}\eta(j)\right]\left[\sum_{j=1}^{t}S(j)\right]\right. \\ &\left. + (I(0) - I_c)\sum_{j=1}^{t}\eta(j)\right\} + C_P\sum_{t=1}^{N}\mathcal{E}[\phi(t)\eta(t)] - C_P P_c\sum_{t=1}^{N}\mathcal{E}\eta(t) \\ &+ C_{PA}\sum_{t=1}^{N}\mathcal{E}[\phi(t)\eta(t) - \phi(t)\eta(t-1) - \phi(t-1)\eta(t) + \phi(t-1)\eta(t-1)] \\ &= 0,\end{aligned} \tag{9}$$

for all admissible $\{\eta\}$. Simplifying (9), we get:

$$\begin{aligned}\Psi = \sum_{t=1}^{N}\left\{C_I\,\mathcal{E}\left[\left(\sum_{j=1}^{t}\phi(j) - \sum_{j=1}^{t}S(j) + I(0) - I_c\right)\left(\sum_{j=1}^{t}\eta(j)\right)\right]\right. \\ + C_P\,\mathcal{E}[\phi(t)\eta(t)] - C_P P_c\,\mathcal{E}\eta(t) \\ \left. + C_{PA}\,\mathcal{E}[(\phi(t) - \phi(t-1))(\eta(t) - \eta(t-1))]\right\} = 0\end{aligned} \tag{10}$$

for all admissible $\{\eta\}$.

Now any set of functions $\{\eta\}$ that are constants, independent of the S's, will be admissible functions. In particular, we let:

$$\eta_k = 1, \qquad \eta_j = 0 \text{ for } j \neq k. \tag{11}$$

Then (10) simplifies to:

$$\begin{aligned}\sum_{t=k}^{N}\left\{C_I \mathcal{E}\left[\sum_{j=1}^{t}\phi(j) - \sum_{j=1}^{t}S(j) + I(0) - I_c\right]\right\} \\ + C_P\,\mathcal{E}[\phi(k)] - C_P P_c \\ + C_{PA}\,\mathcal{E}[\phi(k) - \phi(k-1)] - C_{PA}\,\mathcal{E}[\phi(k+1) - \phi(k)] = 0, \\ (k = 1, \cdots, N),\end{aligned} \tag{12}$$

where, for symmetry, we define $\phi(N + 1) = 0$.

This is a set of N linear equations in $\{\mathcal{E}\phi(j)\}$ and $\{\mathcal{E}S(j)\}$, which can be solved for the former in terms of the latter. Hence, the expected values of the optimal P's are functions only of the expected values of the S's. Since $\phi(1)$ is not a function of any S, we have $\phi(1) = \mathcal{E}\phi(1)$. But $\phi(1)$, which we have now determined to be a function only of the $\{\mathcal{E}S(j)\}$, is the optimal production for the first period. This completes our proof that the certainty-equivalent method leads to the same action as the general programming method, and hence that the former is an optimal decision procedure when the joint probability distribution of future sales is known.

A careful examination of equations (2), (3), and (12) will show in what direction our result can be generalized to a broader class of criterion functions. First, we observe that the kth equation of (12) can be obtained, formally, by setting equal to zero the partial derivative with respect to $P(k)$ of (2). Now our proof requires that the S's appear only linearly in (12), and that there be no cross-products of S's by ϕ's in the terms of that equation. This implies, in turn, that in the expansion of the final form of (2) no term higher than a quadratic appear in products of the S's by P's (e.g., no term of the form $P(j)S^2(k)$). By (3), this implies further that in the first form of (2) no term higher than a quadratic appear in the I's or in products of the I's by P's.

Finally, we observe again that the proof makes no assumption regarding the stationarity or independence of the distribution functions for sales for successive periods. Further, future sales could be made dependent on other stochastic variables in addition to past sales; and the optimal policy could be made to depend on these same variables.

Carnegie Institute of Technology

REFERENCES

[1] Arrow, Kenneth J., Theodore Harris, and Jacob Marschak: "Optimal Inventory Policy," *Econometrica*, 19: 250–272 (July, 1951).

[2] Bellman, Richard: *An Introduction to the Theory of Dynamic Programming*. Santa Monica: The RAND Corporation, 1953.

[3] DVORETSKY, A., J. KIEFER, AND J. WOLFOWITZ: "The Inventory Problem: I," *Econometrica*, 20: 187–222 (April, 1952).

[4] ———: "The Inventory Problem: II," *Econometrica*, 20: 450–466 (July, 1952).

[5] HOLT, CHARLES C. AND HERBERT A. SIMON: "Optimal Decision Rules for Production and Inventory Control," *Proceedings of the Conference on Operations Research in Production and Inventory Control*, January 20–22, 1954. pp. 73–89. Cleveland: Case Institute of Technology, 1954.

[6] HOLT, CHARLES C., FRANCO MODIGLIANI, AND JOHN F. MUTH: "Deriviation and Computation of Linear Decision Rules for Production and Employment Scheduling," to be published in *Management Science*, January, 1956.

[7] HOLT, CHARLES C., FRANCO MODIGLIANI, AND HERBERT A. SIMON: "A Linear Decision Rule for Production and Employment Scheduling," to be published in *Management Science*, October, 1955.

[8] SIMON, HERBERT A.: "Some Properties of Optimal Linear Filters," *Quarterly of Applied Mathematics*, 12: 438–440 (January, 1955).

[9] THEIL, H.: "Econometric Models and Welfare Maximisation," *Weltwirtschaftliches Archiv*, 72: 60–83 (1954, № 1).