

THREE DIFFERENT OPERATIONS RESEARCH MODELS FOR THE SAME (s, S) POLICY

OMAR BEN-AYED[†]

benayed@ksu.edu.sa

*Department of Business Administration, College of Administrative Sciences
King Saud University, PO Box 2459, Riyadh 11451, Kingdom of Saudi Arabia*

Abstract. Operations Research techniques are usually presented as distinct models. Difficult as it may often be, achieving linkage between these models could reveal their interdependency and make them easier for the user to understand. In this article three different models, namely Markov Chain, Dynamic Programming, and Markov Sequential Decision Processes, are used to solve an inventory problem based on the periodic review system. We show how the three models converge to the same (s, S) policy and we provide a numerical example to illustrate such a convergence.

Keywords: Markov Chains, Dynamic Programming, Markov Sequential Decision Processes, Periodic Review System, (s, S) policy.

1. Introduction

Operations Research is usually perceived as a set of models each of which is applicable to a specific type of problems. Operations Research textbooks often fail to establish linkage between these models and deal with them as “unrelated” topics. Such linkage is essential to ensure the integrity of Operations Research. The present article aims at linking three different Operations Research models, namely Markov Chain, Dynamic Programming, and Markov Sequential Decision Processes, by applying each of them to the same inventory control problem. The article seeks to explain how a solution is obtained by each of the three models and how the three solutions are equivalent even though they may look quite different.

[†] Requests for reprints should be sent to O. Ben-Ayed, Department of Business Administration, College of Administrative Sciences, King Saud University, PO Box 2459, Riyadh 11451, Kingdom of Saudi Arabia.

2. The Problem and the (s, S) Policy Solution

Let us consider a hypothetical company estimating the distribution of demand D for one of the items it is producing by $P[D = j] = p_j$, for $j \in \{m, \dots, M\}$; where $P[D = j]$ is the probability of having a level of demand equal to j , and p_j the value of such a probability. The demand for any period n can be satisfied by the quantity x_n produced during period n and/or the quantity i_n available in inventory at the beginning of n . A holding cost c_h is incurred for every unit stored from one period to another, and a stockout cost c_u is incurred for every unit unavailable when requested (lost sale). The production cost $c_g(x_n)$, expressed as a function of the quantity produced x_n , is assumed to be zero when x_n equals zero and is concave for $x_n > 0$.

Since no specific inventory policy has been adopted, the management of the company is now interested in developing a process control system whereby reorder decisions are automatically generated according to a production policy δ_n that associates to each inventory level i_n , at the beginning of the period n , a fixed production quantity $\delta_n(i_n)$ chosen from the set of possible production quantities $\{x_n\}$.

Scarf [1] proved the existence, for each period n , of an optimal production policy δ_n^* that brings the inventory level to a target level S_n^* whenever the initial inventory position i_n for the item is lower than (or equal to) a determined value s_n^* . One important feature of our problem is that cost functions, demand distribution, as well as possible levels of initial inventory, are the same for all periods. This implies the existence of a steady state so that for any possible value of initial inventory i corresponds one optimal policy $\delta^*(i)$ independently of the period n . Therefore, our concern is to find that optimal decision policy δ^* that associates to each inventory position i the production quantity $\delta^*(i)$ that minimizes the total production, holding and stockout costs, for an infinite horizon. Such a policy is determined by the two optimal values s^* and S^* of the two variables s and S , respectively:

$$\delta(i) = \begin{cases} S - i & \text{if } i \leq s \\ 0 & \text{if } i > s \end{cases} \quad (1)$$

Further, the values of i can never exceed S (the highest possible level) minus m (the lowest possible demand):

$$i \in \{0, 1, \dots, S - m\} \quad (2)$$

Constraints (1) and (2) implicitly require that:

$$S > s \text{ and } S \geq m \quad (3)$$

Moreover, we assume an inventory capacity restriction of K units:

$$i + \delta(i) \leq K \Rightarrow S \leq K \quad (4)$$

3. The Markov Chain Model

The inventory level I_n at the beginning of each period n is a discrete-time stochastic process whose possible values are $\{0, 1, \dots, S - m\}$, as stated in (2). Since I_n is always equal to I_{n-1} plus production minus sales, its probability distribution depends on the inventory level I_{n-1} and not on the states the stochastic process passed through on the way to I_{n-1} . For all states i and k and all periods n , the probability that the system is in state i at the beginning of period $n-1$ will be in state k at the beginning of period n , does not depend of n , but does so on the specified policy (s, S) . Therefore, the transition probabilities can be written as $P[I_n = k \mid I_{n-1} = i] = q_{ik}^{sS}$.

Let y and z be two natural numbers verifying $0 \leq y \leq M$ and $m \leq z \leq S$, the transition matrix Q^{sS} from the states $i = 0, 1, \dots, s, s+1, \dots, M - y, \dots, S - z, \dots, S - m$ to the states $k = 0, 1, \dots, M - y, \dots, S - z, \dots, S - m$ can be represented as shown below, where $p_j = 0$ for all $j < 0$ (e.g., $p_{m-z} = 0$ if $z > m$) and $\sum_{i=x}^y p_i = 0$ for all $y < x$ (e.g., $\sum_{i=S}^M p_i = 0$ if $S > M$). At optimality, we must have $s < M$ for if we have enough stock to satisfy all the demand of the period, there will be no need to order and incur unnecessary holding cost [2]. However, as we are uncertain whether $M < S$ or $M > S$, we include the two parameters y and z .

$$Q^{sS} = \begin{array}{c} \begin{array}{cccccc} \text{State } 0 & \dots & \text{State } M-y & \dots & \text{State } S-z & \dots & \text{State } S-m \\ \sum_{i=0}^M p_i & \dots & p_{S-M+y} & \dots & p_z & \dots & p_m \\ \sum_{i=S}^M p_i & \dots & p_{S-M+y} & \dots & p_z & \dots & p_m \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=S}^M p_i & \dots & p_{S-M+y} & \dots & p_z & \dots & p_m \\ \sum_{i=s+1}^M p_i & \dots & p_{s+1-M+y} & \dots & p_{s+1-S+z} & \dots & p_{s+1-S+m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=M-y}^M p_i & \dots & p_0 & \dots & p_{M-y-S+z} & \dots & p_{M-y-S+m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=S-z}^M p_i & \dots & p_{S-z-M+y} & \dots & p_0 & \dots & p_{m-z} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=S-m}^M p_i & \dots & p_{S-m-M+y} & \dots & p_{z-m} & \dots & p_0 \end{array} \\ \begin{array}{l} \text{State } 0 \\ \text{State } 1 \\ \vdots \\ \text{State } s \\ \text{State } s+1 \\ \vdots \\ \text{State } M-y \\ \vdots \\ \text{State } S-z \\ \vdots \\ \text{State } S-m \end{array} \end{array}$$

As none of the states in the chain is transient or periodic, and since all of

them communicate with each other, we can conclude that the chain is ergodic [3], [4]. Therefore, there exists a steady-state distribution $\pi^{sS} = [\pi_0^{sS}, \pi_1^{sS}, \dots, \pi_{S-m}^{sS}]$ for the chain that can be calculated by solving the system:

$$\begin{cases} \pi^{sS} Q^{sS} = \pi^{sS} \\ \mathbf{1} \pi^{sS} = 1; \end{cases} \quad \text{where : } \mathbf{1} = [1, 1, \dots, 1] \quad (5)$$

Let us call $g(s, S)$, $h(s, S)$ and $u(s, S)$ the expected per period production, holding and stockout costs, respectively, as functions of reorder point s and target level S :

$$g(s, S) = \sum_{i=0}^s \pi_i^{sS} c_g (S - i) \quad (6)$$

$$\begin{aligned} h(s, S) = c_h \times & \left[\sum_{i=0}^s \pi_i^{sS} \sum_{j=m}^M \max(0, S - j) p_j \right. \\ & \left. + \sum_{i=s+1}^{S-m} \pi_i^{sS} \sum_{j=m}^M \max(0, i - j) p_j \right] \end{aligned} \quad (7)$$

$$\begin{aligned} u(s, S) = c_u \times & \left[\sum_{i=0}^s \pi_i^{sS} \sum_{j=m}^M \max(0, j - S) p_j + \right. \\ & \left. \sum_{i=s+1}^{S-m} \pi_i^{sS} \sum_{j=m}^M \max(0, j - i) p_j \right] \end{aligned} \quad (8)$$

Let $w(s, S)$ be the expected total cost, equal to the sum of the three functions (6)–(8). The optimal values s^* and S^* can be obtained by minimizing $w(s, S) = g(s, S) + h(s, S) + u(s, S)$ subject to (4):

$$\begin{aligned} \min w(s, S) = & g(s, S) + h(s, S) + u(s, S), \\ \text{S.T. } & s < S \text{ and } S \in \{m, m+1, \dots, K\} \end{aligned} \quad (9)$$

4. The Dynamic Programming Model

Let the period n be the phase and the inventory level i_n at the beginning of the period n the state. The process evolves from state i_n to state i_{n+1} as:

$$i_{n+1} = \max(0, i_n + \delta_n(i_n) - j) \quad (10)$$

where j belongs to the set $\{m, \dots, M\}$ and $\delta_n(i_n)$ is the quantity to produce during period n according to the policy δ_n as a function of the initial inventory level i_n . Let $v(i_n, \delta_n(i_n))$ denote the expected total cost of production, holding, and stockout, for any period n having an initial inventory of i_n units and a production of $\delta_n(i_n)$ units:

$$v(i_n, \delta_n(i_n)) = c_g(\delta_n(i_n)) + \sum_{j=m}^M p_j \times \left[c_h \max(0, i_n + \delta_n(i_n) - j) + c_u \max(0, j - i_n - \delta_n(i_n)) \right] \quad (11)$$

The objective is to minimize the expected total cost for the periods 1, 2, \dots , given that the inventory level is initially i_1 . If we denote the objective function by $f_1(i_1)$, we can generate a more general function $f_n(i_n)$ defined as the minimal expected total costs for the periods $n, n+1, \dots$, given that i_n units are initially available in inventory. The recurrence relation between $f_n(i_n)$ and $f_{n+1}(i_{n+1})$ can be expressed as:

$$f_n(i_n) = \min_{\delta_n(i_n)} \left[v(i_n, \delta_n(i_n)) + \sum_{j=m}^M p_j f_{n+1}(i_{n+1}) \right], \quad \text{for } n = 1, 2, \dots \quad (12)$$

However, dynamic programming models require a finite horizon [5], [6] since $f_n(i_n)$ in (12) cannot be computed before $f_{n+1}(i_{n+1})$. This imposes a last period N as the starting point of the recurrence relation. N could be chosen large enough to enable the process to reach a steady state. For the first periods, one optimal policy $\delta^*(i)$ corresponds to any possible value of initial inventory i , independently of the period n . However, the last periods could be different, as they may carry on the effect of the introduction of the “dummy” last period N . The solution of the dynamic program is achieved first by minimizing $v(i_N, \delta_n(i_n))$ to obtain $\delta_N^*(i)$. Then, we use the recursivity in (12) to find $\delta_{N-1}^*(i)$, $\delta_{N-2}^*(i)$, \dots and so on until the procedure reaches a period $N - L$ verifying $\delta_{N-L}^*(i) = \delta_{N-L-1}^*(i) = \delta_{N-L-2}^*(i) = \dots = \delta_1^*(i) = \delta^*(i)$, thereby solving the problem for the last L periods only:

$$f_N(i) = \min_{\delta_N(i)} \left\{ v(i, \delta_N(i)) \right\} \quad (13)$$

$$f_n(i) = \min_{\delta_n(i)} \left\{ v(i, \delta_n(i)) + \sum_{j=m}^M p_j f_{n+1} \left(\max(0, i + \delta_n(i) - j) \right) \right\},$$

$$\text{for } n = N-1, \dots, N-L \quad (14)$$

$$f_n(i) = \min_{\delta(i)} \left\{ v(i, \delta(i)) + \sum_{j=m}^M p_j f_{n+1} \left(\max(0, i + \delta(i) - j) \right) \right\},$$

$$\text{for } n = N-L-1, \dots, 1 \quad (15)$$

$$i \in \{0, 1, \dots, K-m\} \text{ and } \delta(i), \delta_n(i), \delta_N(i) \in \{0, 1, \dots, K-i\} \quad (16)$$

The first part of (16) is justified exactly in the same way as (2): i can never exceed the highest possible level (K) minus the lowest possible demand (m). The second part is directly obtained from (4).

5. The Markov Sequential Decision Processes Model

A Markov sequential decision process can be defined as an infinite horizon probabilistic dynamic program. It can also be defined as a Markov process with a finite number of states and with an economic value structure associated with the transitions from one state to another [3], [7]. In our case, the state will continue to be the initial inventory of the period. Let $f_\delta(i)$ be the expected cost incurred during an infinite number of periods, given that, at the beginning of period 1, the state is i and stationary policy δ is followed:

$$f_\delta(i) = v(i, \delta(i)) + \sum_{j=m}^M p_j f_\delta \left(\max(0, i + \delta(i) - j) \right) \quad (17)$$

where $v(i, \delta(i))$ is the expected cost incurred during the current period, as defined in (10). The horizon being infinite, $f_\delta(i)$ will also be infinite. To cope with the problem, we can use the expected *discounted* total cost. We assume that a \$1 paid the next period will have the same value as a cost of β dollars paid during the current period. Let $V_\delta(i)$ be the expected discounted cost incurred during an infinite number of periods, given that, at the beginning of period 1, the state is i and stationary policy δ is followed:

$$V_\delta(i) = v(i, \delta(i)) + \beta \sum_{j=m}^M p_j V_\delta \left(\max(0, i + \delta(i) - j) \right) \quad (18)$$

where $\sum_{j=m}^M p_j V_\delta \left(\max(0, i + \delta(i) - j) \right)$ is the expected cost, discounted back to the beginning of period 2 and incurred from the beginning of period 2 onward. The smallest value of $V_\delta(i)$, that we denote by $V(i)$, is the

expected discounted cost incurred during an infinite number of periods, provided that the state at the beginning of period 1 is i and the optimal stationary policy δ^* is followed:

$$V(i) = V_{\delta^*}(i) = \min_{\delta} V_{\delta}(i) \quad \text{for all possible values of } i \quad (19)$$

Using (16) and (18), equality (19) can be equivalently written as:

$$\text{For } i = 0, \dots, K-m: \quad (20)$$

$$V(i) = \min_{\delta(i)=0, \dots, K-i} \left\{ v(i, \delta(i)) + \beta \sum_{j=m}^M p_j V(\max(0, i + \delta(i) - j)) \right\}$$

This can be transformed into the following $K-m$ linear programs:

$$\max V(i); \quad \text{for } i = 0, \dots, K-m \quad (21)$$

$$\text{S.T.} \quad (22)$$

$$V(i) \leq v(i, \delta(i)) + \beta \sum_{j=m}^M p_j V(\max(0, i + \delta(i) - j))$$

$$\delta(i) = 0, \dots, K-i$$

It can be shown [8] that the solutions of the K inter-dependent linear programs (21)–(22) are achieved simply by taking the sum of all the objectives, thus obtaining a single-objective linear program:

$$\max \sum_{i=0}^{K-m} V(i) \quad (23)$$

$$\text{S.T.} \quad (24)$$

$$V(i) \leq v(i, \delta(i)) + \beta \sum_{j=m}^M p_j V(\max(0, i + \delta(i) - j))$$

$$i=0, \dots, K-m; \quad \delta(i)=0, \dots, K-i$$

6. Linking the Models

First we show the link between the last two models, then between the first and the last ones.

6.1. Linking the Dynamic Programming Model and the Markov Decision Process Model

The solution of the dynamic program is that of (15)–(16). However, as the horizon is initially infinite, we can choose n sufficiently large so that $f_n(i) = f_{n+1}(i) = f(i)$. This allows the writing of (15)–(16) as:

$$\begin{aligned} &\text{For } i = 0, \dots, K-m: & (25) \\ f(i) &= \min_{\delta(i)=0, \dots, K-i} \left\{ v(i, \delta(i)) + \sum_{j=m}^M p_j f(\max(0, i + \delta(i) - j)) \right\} \end{aligned}$$

The same equality can be obtained when giving β the value of 1 in (20):

$$\begin{aligned} &\text{For } i = 0, \dots, K-m: & (26) \\ V(i) &= \min_{\delta(i)=0, \dots, K-i} \left\{ v(i, \delta(i)) + \sum_{j=m}^M p_j V(\max(0, i + \delta(i) - j)) \right\} \end{aligned}$$

Therefore, both (15)–(16) and (20) are obtained from (25). The two models diverged when dealing with the problem of the infinite value of the function (25). In (15)–(16) a finite number of periods was fixed and in (20) the expected cost was discounted.

6.2. Linking the Markov Decision Process Model and the Markov Chain Model

Let us focus on (17), which was the starting point of the Markov sequential decision processes model. To simplify the representation, we assume that the state evolves from i_0 to i_{j_0} , then $i_{j_1}, i_{j_2}, i_{j_3} \dots$. This means that we denote $\max(0, i + \delta(i) - j_k)$ by i_{j_k} :

$$\begin{aligned} f_\delta(i_0) &= v(i_0, \delta(i_0)) + \sum_{j_0=m}^M p_{j_0} f_\delta(i_{j_0}) \\ &= v(i_0, \delta(i_0)) + \sum_{j_0=m}^M p_{j_0} \left[v(i_{j_0}, \delta(i_{j_0})) + \sum_{j_1=m}^M p_{j_1} f_\delta(i_{j_1}) \right] \\ &= v(i_0, \delta(i_0)) + \sum_{j_0=m}^M p_{j_0} v(i_{j_0}, \delta(i_{j_0})) + \sum_{j_0=m}^M p_{j_0} \sum_{j_1=m}^M p_{j_1} v(i_{j_1}, \delta(i_{j_1})) \end{aligned}$$

$$\begin{aligned}
& + \dots + \sum_{j_0=m}^M p_{j_0} \sum_{j_1=m}^M p_{j_1} \sum_{j_2=m}^M p_{j_2} \dots \sum_{j_k=m}^M p_{j_k} v(i_{j_k}, \delta(i_{j_k})) \\
& + \sum_{j_0=m}^M p_{j_0} \sum_{j_1=m}^M p_{j_1} \sum_{j_2=m}^M p_{j_2} \dots \sum_{j_k=m}^M p_{j_k} \sum_{j_{k+1}=m}^M p_{j_{k+1}} f_\delta(i_{j_{k+1}}) \quad (27)
\end{aligned}$$

There is no end to the sequence $\{i_{j_0}, i_{j_1}, \dots, i_{j_k}, \dots\}$. However, as stated in (16), the possible values of $i_{j_0}, i_{j_1}, \dots, i_{j_k}, \dots$ are finite and belong to $\{0, 1, \dots, K-m\}$, which can be interpreted as:

$$\begin{aligned}
v(i_{j_k}, \delta(i_{j_k})) & \in \left\{ v(0, \delta(0)), v(1, \delta(1)) \dots, v(K-m, \delta(K-m)) \right\}, \\
k & = 0, 1, 2, \dots \quad (28)
\end{aligned}$$

The frequency of occurrence of $v(i, \delta(i))$ in (27) varies from one strategy δ to another. When denoting such a frequency by N_i^δ , we can combine (27) and (28) as:

$$f_\delta(i_0) = \sum_{i=0}^{K-m} N_i^\delta v(i, \delta(i)) = f_\delta \quad (29)$$

f_δ in (29), which is the same as $f_\delta(i)$ in (17), is infinite because N_i^δ 's are infinite. To cope with the problem, we can take the average cost per period that we denote by \bar{f}_δ (instead of the total cost for the whole horizon f_δ). Let us denote by π_i^δ the relative frequency of incurring the cost $v(i, \delta(i))$ when policy δ is followed. Using (29) and (1), we can write:

$$\begin{aligned}
\bar{f}_\delta & = \frac{f_\delta}{\sum_{i=0}^{K-m} N_i^\delta} = \sum_{i=0}^{K-m} \frac{N_i^\delta}{\sum_{i=0}^{K-m} N_i^\delta} v(i, \delta(i)) = \sum_{i=0}^{K-m} \pi_i^\delta v(i, \delta(i)) \\
& = \sum_{i=0}^{K-m} \pi_i^\delta \left\{ c_g(\delta(i)) + \sum_{j=m}^M p_j \times \left[c_h \max(0, i + \delta(i) - j) + \right. \right. \\
& \quad \left. \left. c_u \max(0, j - i - \delta(i)) \right] \right\} \\
& = \sum_{i=0}^s \pi_i^\delta c_g(\delta(i)) + \sum_{i=0}^s \pi_i^\delta \sum_{j=m}^M p_j c_h \max(0, i + \delta(i) - j) \\
& \quad + \sum_{i=s+1}^{S-m} \pi_i^\delta \sum_{j=m}^M p_j c_h \max(0, i + \delta(i) - j)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i=0}^s \pi_i^\delta \sum_{j=m}^M p_j c_u \max(0, j-i-\delta(i)) \\
& + \sum_{i=s+1}^{S-m} \pi_i^\delta \sum_{j=m}^M p_j c_u \max(0, j-i-\delta(i)) + \sum_{i=s+1}^{K-m} \pi_i^\delta c_g(\delta(i)) \\
& + \sum_{i=S-m+1}^{K-m} \pi_i^\delta \sum_{j=m}^M p_j c_h \max(0, i+\delta(i)-j) \\
& + \sum_{i=S-m+1}^{K-m} \pi_i^\delta \sum_{j=m}^M p_j c_u \max(0, j-i-\delta(i)) \\
& = \sum_{i=0}^s \pi_i^\delta c_g(S-i) + \sum_{i=0}^s \pi_i^\delta \sum_{j=m}^M p_j c_h \max(0, S-j) \\
& + \sum_{i=s+1}^{S-m} \pi_i^\delta \sum_{j=m}^M p_j c_h \max(0, i-j) \\
& + \sum_{i=0}^s \pi_i^\delta \sum_{j=m}^M p_j c_u \max(0, j-S) + \sum_{i=s+1}^{S-m} \pi_i^\delta \sum_{j=m}^M p_j c_u \max(0, j-i) \\
& = g(s, S) + h(s, S) + u(s, S) = w(s, S) \tag{30}
\end{aligned}$$

In other words, the expected total cost $w(s, S)$ in the Markov chain model (8) is in fact the average cost per period obtained from the expected cost (17) in the Markov sequential decision process model, using equality (1) from which the constraints of (8) were derived. Both models were based on the infinite function (17). They diverged when dealing with infinity; the Markov sequential decision process model used the expected discounted cost while the Markov chain model used the average cost per period.

7. Numerical Application

Assume that demand is either 1 or 3 units with respective probabilities $p_1 = \frac{1}{3}$ and $p_3 = \frac{2}{3}$, unit holding and stockout costs are $c_h = \$5$ and $c_u = \$8$, production costs as a function of the possible values are $c_g(0) = 0$, $c_g(1) = 10$, $c_g(2) = 16$, and $c_g(3) = 18$. Accordingly, we can write: $m = 1$ and $M = K = 3$, which means that $S \in \{1, 2, 3\}$ (as $m \leq S \leq K$) and $s \in \{0, 1, 2\}$ (as $s < S$).

7.1. Solution of the Markov Chain Model

Based on the possible values of S and s , we have to choose one among six possible policies: δ^{01} , δ^{02} , δ^{03} , δ^{12} , δ^{13} , and δ^{23} , where δ^{sS} denotes the policy (s, S) . The corresponding Q^{sS} matrices (Q^{01} , Q^{02} , Q^{03} , Q^{12} , Q^{13} and Q^{23}) will be:

$$Q^{01}=[1]; \quad Q^{02}=\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ 1 & 0 \end{bmatrix}; \quad Q^{03}=\begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{3} \\ 1 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}; \quad Q^{12}=\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

$$Q^{13}=\begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{bmatrix}; \quad Q^{23}=\begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{3} \\ \frac{2}{3} & 0 & \frac{1}{3} \\ 3 & 0 & \frac{1}{3} \end{bmatrix}$$

We apply (5) to get the steady state probabilities π^{sS} for each (s, S) policy:

$$\pi^{01}=[1]; \quad \pi^{02}=\begin{bmatrix} \frac{3}{4} & \frac{1}{4} \end{bmatrix}; \quad \pi^{03}=\begin{bmatrix} \frac{9}{13} & \frac{1}{13} & \frac{3}{13} \end{bmatrix}; \quad \pi^{12}=\begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}; \quad \pi^{13}=\begin{bmatrix} \frac{2}{3} & \frac{1}{12} & \frac{1}{4} \end{bmatrix}; \quad \pi^{23}=\begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{3} \end{bmatrix}$$

The corresponding expected total costs $w(s, S)$, as defined in (8), are $w(0, 1) = \frac{62}{3}$; $w(0, 2) = \frac{239}{12}$; $w(0, 3) = \frac{671}{39}$; $w(1, 2) = 21$; $w(1, 3) = \frac{211}{12}$; and $w(2, 3) = \frac{62}{3}$. The lowest value being $\frac{671}{39}$, we conclude that $(0, 3)$ is the best policy.

7.2. Solution of the Dynamic Programming Model

The solutions for the periods N and $N-1$ are provided in the following table where i , i_{N-1} , i_N , $\delta(i)$, $\delta_{N-1}(i)$ and $\delta_N(i)$ are as defined in (16), $v(i, \delta(i))$ as defined in (10) and i_N is $\max(0, i + \delta(i) - j)$ as defined in (10). Based on the last column of the table, the optimal policy is to produce 3 only when $i=0$. The same solution is obtained for f_{N-2} , f_{N-3} , ... (calculations not shown), which means that $(0,3)$ is the optimal steady state policy (as found previously).

i	$v(i, \delta_N(i))$, for $\delta_N(i) =$				$f_N(i)$	$\delta_N^*(i)$	$v(i, \delta_{N-1}(i)) +$, $\sum_{j=m}^M p_j f_N(i_N)$, for $\delta_{N-1}(i) =$				$f_{N-1}(i)$	$\delta_{N-1}^*(i)$
	0	1	2	3			0	1	2	3		
0	$\frac{56}{3}$	$\frac{62}{3}$	23	$\frac{64}{3}$	$\frac{56}{3}$	0	$\frac{112}{3}$	$\frac{118}{3}$	39	$\frac{325}{9}$	$\frac{325}{9}$	3
1	$\frac{32}{3}$	17	$\frac{58}{3}$	-	$\frac{32}{3}$	0	$\frac{88}{3}$	33	$\frac{307}{9}$	$\frac{88}{3}$	-	0
2	7	$\frac{40}{3}$	-	-	7	0	23	$\frac{253}{3}$	-	23	-	0

If we use the steady state probabilities π^{03} computed earlier, we can find the same expected total cost per period: $\sum_{i=0}^2 \pi_i^{03} v(i, \delta^*(i)) = \frac{9}{13} \times \frac{64}{3} + \frac{1}{13} \times \frac{32}{3} + \frac{3}{13} \times 7 = \frac{671}{39} = w(0, 3)$.

7.3. Solution of the Markov Sequential Decision Processes Model

Assuming $\beta = .985$, the following Linear Program is obtained by applying (23)–(24):

$$\begin{aligned} & \max V(0) + V(1) + V(2) \\ \text{S.T. } & V(0) \leq v(0, 0) + .985 \left[\frac{1}{3} V(0) + \frac{2}{3} V(0) \right]; v(0, 0) = \frac{56}{3} \\ & V(0) \leq v(0, 1) + .985 \left[\frac{1}{3} V(0) + \frac{2}{3} V(0) \right]; v(0, 1) = \frac{62}{3} \\ & V(0) \leq v(0, 2) + .985 \left[\frac{1}{3} V(1) + \frac{2}{3} V(0) \right]; v(0, 2) = 23 \\ & V(0) \leq v(0, 3) + .985 \left[\frac{1}{3} V(2) + \frac{2}{3} V(0) \right]; v(0, 3) = \frac{64}{3} \\ & V(1) \leq v(1, 0) + .985 \left[\frac{1}{3} V(0) + \frac{2}{3} V(0) \right]; v(1, 0) = \frac{32}{3} \\ & V(1) \leq v(1, 1) + .985 \left[\frac{1}{3} V(1) + \frac{2}{3} V(0) \right]; v(1, 1) = 17 \\ & V(1) \leq v(1, 2) + .985 \left[\frac{1}{3} V(2) + \frac{2}{3} V(0) \right]; v(2, 0) = \frac{58}{3} \\ & V(2) \leq v(2, 0) + .985 \left[\frac{1}{3} V(1) + \frac{2}{3} V(0) \right]; v(2, 0) = 7 \\ & V(2) \leq v(2, 1) + .985 \left[\frac{1}{3} V(1) + \frac{2}{3} V(0) \right]; v(2, 1) = \frac{40}{3} \end{aligned}$$

which leads to the solution $V(0) = 1150.382$, $V(1) = 1143.793$, and $V(2) = 1137.963$. The discounted expected cost for the infinite horizon, that we denote by W , can be calculated on the basis of the steady state probabilities:

$$W = \frac{9}{13} \times 1150.382 + \frac{1}{13} \times 1143.793 + \frac{3}{13} \times 1137.963 = 11147.010$$

The same value of W could be found by dividing $w(0, 3)$ by $1 - \beta$:

$$W = 11147.010 = \frac{\frac{671}{39}}{1 - .985} = \frac{w(0, 3)}{1 - \beta}$$

This illustrates the convergence of the three models.

8. Conclusion

In this paper, we used three different models to solve the same problem based on the same notation, the same data, and the same assumptions. Despite some similarities, the three models approached the problem in different ways. Having different theoretical bases, the obtained formulations showed major differences, but they all converged into the same optimal solution as was illustrated by the numerical application. Such a convergence

is justified by the fact that all three models lead to an exact solution, which is the optimal (s, S) policy.

Acknowledgments

The author wishes to thank the anonymous referee for his efforts in shaping the manuscript and for his helpful comments and suggestions which improved the content of the article.

References

1. H. Scarf, "The optimality of (s, S) policies for the Dynamic Inventory Problem", *Proceedings of the First Stanford Symposium on Mathematical Methods in the Social Sciences*, Stanford University Press, 1960.
2. H. Wagner and T. Whitin, "Dynamic Version of the Economic Lot Size Model", *Management Science*, **5**, 1, 1958.
3. D. Isaacson and R. Madsen, *Markov Chains: Theory and Applications*, John Wiley, 1975.
4. W. L. Winston, *Operations Research: Applications and Algorithms*, PWS-KENT, 1994, Third Edition.
5. L. Cooper and M.W. Cooper, *Introduction to Dynamic Programming*, Pergamon Press, 1981.
6. D. Bertsekas, *Dynamic Programming*, Prentice-Hall, 1987.
7. S. Kohlas, *Stochastic Methods of Operations Research*, Cambridge University Press, 1982.
8. S. Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, 1983.