IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 16, Issue 3 Ser. I (May – June 2020), PP 08-17
www.iosrjournals.org

An Optimal Ordering Policy with Markov Decision Process

Aisha Sheikh Hassan,YahayaAhmad Abubakar, Usman Abdullahi M, Philip Moses Auduand Bello Yakubu
Mathematics Department, Federal Polytechnic Bida, Niger State, Nigeria.

ABSTRACT: *One of the most frequent decisions faced by operations managers is "how much" or "how many" items are they to make or buy in order to satisfy external or internal requirements for the item. Replenishment in many cases is made using the economic order quantity (EOQ) model. The model considers the tradeoff between ordering cost and storage cost in choosing the quantity to use in replenishing items in inventories. This paper demonstrates an approach to optimize the EOQ of an item under a periodic review inventory system with stochastic demand using value iteration. The objective is to determine in each period of the planning horizon, an optimal decision so that the long run costs are minimized and profits are maximized for the given state of demands. Using Markov decision process over a finite planning horizon with equal intervals, the decision of how much quantity to order or not to order is made. We use a numerical example with the aid of value iteration method to demonstrate the existence of an optimal decision policy.*

KEYWORDS: *Markov decision process, inventory management, optimization, EOQ, Markov chain, stochastic process, value iteration.*

Date of Submission: 29-04-2020 Date of Acceptance: 13-05-2020

I. Introduction

Inventory management plays a very important role in supply chain. To manufacturers, it entails managing product stocks, in-process inventories of intermediate products as well as inventories of raw material, equipment and tools, spare parts, supplies used in production, and general maintenance supplies. A manufacturing company needs an inventory policy for each of its products to govern when and how much it should be replenished. Good inventory management offers the potential not only to cut costs but also to generate new revenues and higher profits. On the other hand, undersupply causes stock out and leads to lost sales; whereas oversupply hinders free cash flow and may cause forced markdowns. As a result of improper inventory policies, both will diminish earnings and can have enough impact to make a company non-profitable. Due to the ever changing market conditions, the dynamic and random nature of the demands, the close and complicated relationship between resource/production planning and product inventory management, as well as the process uncertainties, matching supply with demand has always been a great challenge. Being able to offer the right product at the right time for the right price remains frustratingly elusive to manufacturers and retailers (Fisher et al 2000). Process scheduling and planning have attracted growing attention in many industriesThe main objective of inventory management is to increase profitability. A frequently used criterion for choosing the optimal policy is to minimize the total costs, which is equivalent to maximizing the net income in many cases. Scientific inventory management requires a sound mathematical model to describe the behavior of the underlying system and, quite often, an optimal policy with respect to the model. Many models have been developed for various inventory situations. The first inventory model appeared in the literature more than 70 years ago (Wilson, 1934) is frequently referred to as the Wilson formulation. This is a fixed order quantity system that selects the order quantity to minimize the total costs in the inventory management. Several of its variations, such as modified reorder point system with periodic inventory counts, the replenishment system, and multiple reorder systems etc. have been widely used. Many inventory systems possess complications that require models capable of handling specific problems in certain situations. Despite the large number of models developed, however, there is still a wide gap between theory and practice. Similar to many other dynamic processes in the real world, demand variation encountered by retailers or manufacturers is both random and seasonal in nature. A random/stochastic process may be considered as an ensemble of random variables defined on a common probability space and evolving over time. The observed data are statistical time series, which are single realizations of the underlying process. Contrary to those from the deterministic processes, the outcomes from a stochastic process is not unique. Time series of on-line data collected from repetitions of the same experiment will not be the same; levels of demand for a product change from week to week. It is desirable or sometimes necessary to quantify the dynamic relationships among these random events so as to better understand and effectively handle process uncertainties. Considering that the dynamics of such systems are

often governed by Markov chains, we resort to Markovian models for solution. Markov chain, a well-known subject introduced by Andrei A. Markov in 1906, has been studied by a host of researchers for many years. Markovian formulations (as in Chiang, 1980; Taylor &Karlin, 1998; Yang et al 2002; Yin et al 2001; Yin & Zhang, 1997, 1998; Yin et al 1995) are useful in solving a number of real-world problems under uncertainties such as determining the inventory levels for retailers, maintenance scheduling for manufacturers, and scheduling and planning in production management. Markov chain approach has been applied in the design, optimization, and control of queueing systems, manufacturing processes, reliability studies and communication networks, where the underlying system is formulated as stochastic control problem driven by Markovian noise. Markov decision process offers an elegant mathematical framework for addressing arbitrarily challenging, sequential decision problems that arise in the fields of operations research, management science, finance, and computer science, among others. Fundamentally, Markov decision process enable researchers to analyze the dynamics of a stochastic process whose transition mechanism is controlled over time: The state of the process provides the decision maker with all the information necessary to choose a feasible action in that state. The process responds to the chosen action by randomly evolving to a new state, and yields either costs or rewards to the decision maker. While Markov decision process captures complex systems, they still enable clean analytical formulations with the help of abstraction and assumptions. Most importantly, it is assumed that the probability that the controlled process transitions into its new state depends only on the current state and the chosen action. In other words, the state transitions of a Markov decision process possess the memoryless property, which greatly simplifies the analysis of stochastic processes. Due to the memoryless assumption, in a Markov decision process one needs to make decisions only at certain time epochs. Therefore, the strength of Markov decision process lies in their ability to be used to formulate a discrete recursive value function capturing the expected cost or reward; the optimal action as a function of the current state can be derived by calculating this value function.

Many researchers have studied various techniques in this context, including dynamic and linear programming, to compute value functions. However, most computational methods suffer from multiple dimensionality; their practical applications are limited to cases where the state space is manageably small and/or the value function has a simple analytical form. To solve computationally nontrivial problems, many other researchers have focused on characterizing the structural properties of value functions. Establishing basic properties of value functions in Markov decision process and showing that they survive under iteration, forms the basis of the inductive proof technique. This technique allows the structure of the optimal policy to be deduced. Structural properties provide a powerful methodology for either partial or complete characterization of optimal policies, which might have important managerial implications and/or offer smarter computational methods. Putterman (1994)

In this work, an optimization model is developed for determining the EOQ under a periodic review inventory system with stochastic demand using value iteration. Adopting a Markov decision approach, the states of the Markov chain represent possible states of demand for the models. The aim is to determine in each period of the planning horizon, an optimal decisionpolicy so that the long run profits are maximized, costs minimized for a given state of demands. Using equal intervals, the decision of how much quantity to order or not to order are made using Markov decision process over a finite planning horizon. A numerical example with the aid of value iteration demonstrates the existence of an optimal decision policy.

Zheng (1992) analyzed a stochastic order quantity and reorder point model in comparison with a corresponding deterministic EOQ model. The research result indicated that at large quantities, the difference between deterministic and stochastic models is small and the relative increase of the cost incurred by using the quantity determined by the EOQ instead of the optimal from the stochastic model does not exceed one eighth and vanishes when ordering costs are significant relative to other costs. Cheung and Powell (1996), formulated a two stage model that minimized the cost of stochastic demand. The first stage dealt with moving inventory from the plant to the warehouses based on forecasted demand. The second stage was moving the inventory from the warehouses to the customers when they send an order. Using an experimental case, the model indicated that having two warehouses per customer was more efficient than having one warehouse per customer.

Eynan and Kropp (1998) examined a periodic review system under stochastic demand using a single product. A simple solution procedure gave an almost optimal solution where results were extended to the joint replenishment problem for multiple items and the simple heuristic developed provided promising results.Piperagkas et al. (2012) investigated the dynamic lot-size problem under stochastic and non-stationary demand over the planning horizon. The problem is solving by three popular meta-heuristic methods from the fields of evolutionary computation and swarm intelligence. Yin et al. (2002) proposed a formulation and solution procedure for inventory planning with the Markov decision process (MDP) models. They formulated the Markov decision model by identifying the chain's state space and the transition probabilities, specify the cost structure and evaluate its individual component; and then use the policy-improvement algorithm to obtain the optimal policy.Broekmeullen and VanDonselarr (2006) developed a replenishment inventory model to understand product, sales and supply characteristics of perishables in supermarkets, analyzed a perishable

inventory control system based on item agingand retrieval behavior, investigates how the intelligence in automated store ordering systems in supermarkets can be further improved and had profound insights in terms of randomdemand. Roychowdhury (2009) determined an optimal policy for a stochastic inventory model of deteriorating items with time dependent selling price. The rate of deterioration of the items was constant over time and the selling price decreased monotonically at a constant rate with deterioration of items. Mubiru and Buhwezi(2017) considered a joint location inventory replenishment problem involving a chain of supermarkets at designated locations. Associated with each supermarket is stochastic stationary demand where inventory replenishment periods are uniformly fixed for the supermarkets. Considering inventory positions of the supermarket chain, they formulated a finite state Markov decision process model where states of a Markov chain represent possible states of demand for milk powder product. The unit replenishment cost, shortage cost, demand and inventory positions were used to generate the total inventory cost matrix; representing the long run measure of performance for the Markov decision process problem. The problem was to determine for each supermarket at a specific location an optimal replenishment policy so that the long run inventory costs are minimized for the given states of demand.Mubiru et al (2019) considered an internet cafe faced with an optimal choice of bandwidthfor internet users under stochastic stationary demand. The choice was made overuniformly time horizons with the goal of optimizing profits. Considering customerdemand, price and operating costs of internet service, they formulated a finite stateMarkov decision process model where states of a Markov chain represented possiblestates of demand for internet service. A profit matrix was generated, representing thelong run measure of performance for the Markov decision process problem.

Kallen and van Noortwijk (2006) presented a decision model for determining the optimal time between periodic inspections of an object with sequential discrete states. The deterioration model used a Markov process to model the uncertain rate of transitioning from one state to the next, allowing the decision maker to properly propagate the uncertainty of the component's condition over time. The model was illustrated by an application to the periodic inspection of road bridges. The author also showed that the model could be applied to production facilities to optimize the threshold for preventive maintenance.

Saranga and Knezevic (2001) developed a mathematical model for reliability prediction of condition-based maintained systems in which the component deterioration was modeled as a Markov process. A system of integral equations was used to compute the reliability of the system at any instant of operating time. When the reliability of the item reached the minimum required reliability level, it was assumed that the item has reached a critical state and hence the required maintenance activities should be carried out to restore the system to an acceptable level. The authors suggested that a well-designed condition monitoring strategy incorporated into condition based maintenance (CBM) could offer improved reliability and availability at the system level.

Sloan and Shanthikumar (2002) considered the problem of determining the production and maintenance schedules for a multiple-product, multiple-stage production system. Each stage consisted of a machine whose condition deteriorated over time and the condition affected the yield of different product types differently. The authors developed a Markov decision process model to simultaneously determine the equipment maintenance and production schedules for each stage of the system with the objective of maximizing the long-run expected average profit. A simulation model of a four-station semiconductor wafer fab was used to compare the performance of policies generated by their model against a variety of other maintenance and dispatching policy combinations. The results indicated that their method provided substantial improvements over traditional methods and performed better as the diversity of the product set increased. They showed that the reward earned using the policies from the combined production and maintenance scheduling method was an average of more than 70% higher than the reward earned using other policy combinations such as a fixed-state maintenance policy and a first come, first-serve dispatching policy.

II. Development Of Value Iteration Method

Let X_n denote the state of the process at time n and a_n the action chosen at time n, then the above is equivalent to stating that:

$$P(X_{n+1} = j \mid X_0, a_0, X_1, a_1, ...X_n = i, a_n = a) = P_{ij}(a) \tag{1.1}$$

We consider an aperiodic irreducible Markov chain with m states (m$< \infty$)and the transition probability matrix

$$\mathrm{P} = \begin{pmatrix} p_{11} & p_{12} & .....p_{1m} \\ . & . & . \\ . & . & . \\ p_{m1} & p_{m2} & ..... p_{mm} \end{pmatrix} \tag{1.2}$$

With every transition, i to j associate a reward R_{ij} if we let $V_i^{(n)}$ be the expected total earnings (reward) in the next n transitions, given that the system is in state i at present.
A simple relation can be given

For $\left\{V_i^{(n)}\right\}_{n-1}^{\infty}$ as follows:

$$V_i^{(n)} = \sum_{j=1}^{m} P_{ij}\left[R_{ij} + V_j^{(n-1)}\right] \quad i = 1, 2, ..., m; n = 1, 2, 3, ... \quad (1.3)$$

Let $$\sum_{j=1}^{m} P_{ij} R_{ij} = Q_i \quad (1.4)$$

Equation can now be written as:

$$V_i^{(n)} = Q_i + \sum_{j=1}^{m} P_{ij} V_j^{(n-1)} \quad (1.5)$$

Setting n = 1, 2 … we get

$$V_i^{(1)} = Q_i + \sum_{j=1}^{m} P_{ij} V^{(0)} \quad (1.6)$$

$$V_i^{(2)} = Q_i + \sum_{j=1}^{m} P_{ij}\left[Q_j + \sum_{k=1}^{m} P_{jk} V_k^{(0)}\right] = Q_i + \sum_{j=1}^{m} P_{ij} Q_j + \sum_{k=1}^{m}\sum_{j=1}^{m} P_{ij} P_{jk} V_k^{(0)} \quad (1.7)$$

$$= Q_i + \sum_{j=1}^{m} P_{ij} Q_j + \sum_{k=1}^{m}\sum_{j=1}^{m} P_{ij} P_{jk} V_k^{(0)} = Q_i + \sum_{j=1}^{m} P_{ij} Q_j + \sum_{k=1}^{m} P_{ik}^{(2)} V_k^{(0)} \quad (1.8)$$

Where $P_{ij}^{(n)}$ is the $(i, j)^{th}$ element of the matrix P^n

$$V^{(n)} = \begin{bmatrix} V_1^{(n)} \\ V_2^{(n)} \\ V_3^{(n)} \\ . \\ . \\ V_m^{(n)} \end{bmatrix} \qquad Q = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ . \\ . \\ Q_m \end{bmatrix}$$

Equation (1.8) can be put in matrix notation as

$$V^{(n)} = Q + PQ + P^2 V^{(0)} \quad (1.9)$$

Extending this to a general n, we have

$$V^{(n)} = Q + PQ + P^2 Q + ... + P^{(n-1)} Q + P^n V^{(0)} = \left[1 + \sum_{k-1}^{n-1} PK\right] Q + P^n V^{(0)} \quad (2.0)$$

we consider the transition probability matrix P and the reward matrix R as given. suppose that the decision maker has other alternatives and so is able to alter elements of P and R. to incorporate this feature, we define D as the decision set, and under a decision $k \in D, let\ ^k P_{ij}\ and\ ^k R_{ij}$ be the probability of the transition and corresponding reward, respectively. Let $^k V_i^{(n)}$ be the expected earnings in n transitions $i \to j$ under decision k, we have the recurrence relations ($k = o$ represents the optimal decision)

$$ {}^{o}V_i^{(n)} = \max_{k \in D} \sum_{j=1}^{m} {}^{k}P_{ij} \left[{}^{k}R_{ij} + {}^{o}V_j^{(n-1)} \right] \; n = 1,2,....; i = 1,2,....., m \qquad (2.1) $$

Giving

$$ {}^{o}V_i^{(n)} = \max_{k \in D} \left[{}^{k}Q_i + \sum_{j=1}^{m} {}^{k}P_{ij} \, {}^{o}V_j^{(n-1)} \right] \; n = 1,2,....; i = 1,2,....., m \qquad (2.2) $$

Where $\sum_{j=1}^{m} {}^{k}P_{ij} \, {}^{k}R_{ij} = {}^{k}Q_i$. Recursive relation (2.2) gives an iterative procedure to determine the optimum decisions $d_i^{(n)} \in D$, for $n = 1,2,....; i = 1,2,....., m$

This is the standard technique in dynamic programming and it has been shown Bellman, (1957) that this iteration process will converge on best alternative for each state as $n \to \infty$. The method is based on recursively determining the optimum policy for every n that would give the maximum value. However, one major drawback of the method is that, there is no way to say when the policy converges into a stable policy; therefore, the value iteration procedure is useful only when n is fairly small.

III. Model Development

In formulating the model, an inventory system of a single product is considered. The demand during each time period over a fixed planning horizon is classified under three states: favorable state (f), less favorable state (l) and unfavorable state (u). The transition probabilities over the planning horizon from one demand state to another could be described by means of a Markov decision process, as such the demand during each period is assumed to depend on the demand of the preceding period. To obtain an optimal course of action, a decision alternatives are open to the decision maker, that is to order large quantity of item, order small quantity of item andnot to order additional units has to be made during each period over the planning horizon, where k is a decision variable. The maximum expected earnings are put together at the end of the period to obtain optimality with the aid of value iteration based on the following transition probability and reward matrix.

$$ \mathrm{P} = \begin{pmatrix} p_{11} & p_{12} & .....p_{1m} \\ . & . & . \\ . & . & . \\ p_{m1} & p_{m2} & ..... p_{mm} \end{pmatrix} \qquad (3.1) $$

$$ R = \begin{pmatrix} R_{11} & R_{12} & ....R_{1m} \\ .. & .. & ..... \\ R_{m1} & R_{m2} & ....R_{mm} \end{pmatrix} \qquad (3.2) $$

The transition between the states is described in by the following transition diagram.

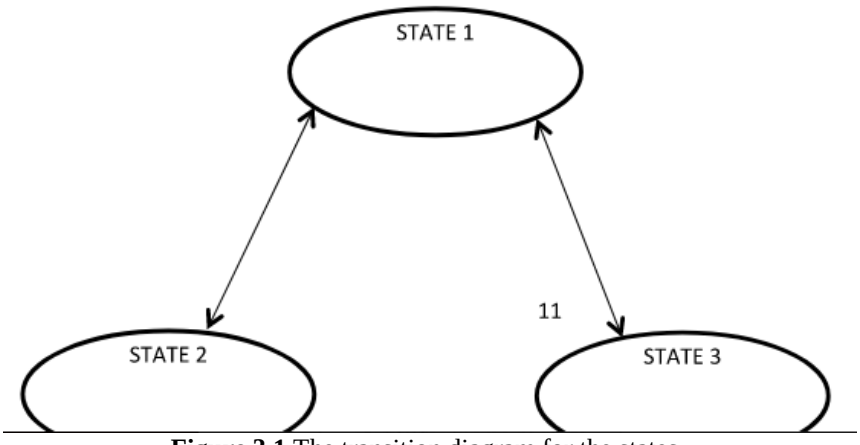

Figure 3.1 The transition diagram for the states.

Let the position of the demands for the product be described by a random variable (X), suppose that the demands is considered for several weeks; (n), we obtain a stochastic process $X_n,\ n=1,2,3,.....$ we assume that the position of the demands are:

(1) Favorable demand (state1)
(2) Less favorable demand (state2)
(3) Unfavorable demand (state3)

We consider the states to be mutually exclusive and exhaustive. It is further assumed that the stochastic process $X_n,\ n=1,2,3,.....$ is governed by a first order Markov chain

$$P_{ij}=P(X_{t+1}=j\,|\,X_0=i_0,X_1=i_1,.....,X_t=i)=P(X_{t+1}=j\,|\,X_t=i). \tag{3.3}$$

The possible transitions between the states are presented in figure (3.1).

From the transition diagram in figure (3.1) and equation (1.1) where m, n = 1, 2. We obtain a transition matrix

$$P=\begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{31} & P_{33} \end{pmatrix} \tag{3.4}$$

We assume that the matrix is P is aperiodic, irreducible stochastic matrix and satisfies equation

$$P_{ij}=P(X_{t+1}=j\,|\,X_0=i_0,X_1=i_1,.....,X_t=i)=P(X_{t+1}=j\,|\,X_t=i) \tag{3.5}$$

When the demand of product is in state 1, two alternatives are open to the decision maker. That is to:

i) Order large quantity
ii) Hold on to the advertisement method.

Let the corresponding transition probabilities and rewards matrix be given as:

$$\begin{pmatrix} {}^1P_{11} & {}^1P_{12} & {}^1P_{13} \\ {}^1P_{21} & {}^1P_{22} & {}^1P_{23} \end{pmatrix} \tag{3.6}$$

$$\begin{pmatrix} {}^1R_{11} & {}^1R_{12} & {}^1R_{13} \\ {}^1R_{21} & {}^1R_{22} & {}^1R_{23} \end{pmatrix} \tag{3.7}$$

When the demand of the product is in state 2, two alternatives are open to the decision maker, that is to:

i) Order small quantity

ii) Increase advertisement

let the corresponding transition probability and reward matrix be given as:

$$\begin{pmatrix} {}^2P_{11} & {}^2P_{12} & {}^2P_{13} \\ {}^2P_{21} & {}^2P_{22} & {}^2P_{23} \end{pmatrix} \tag{3.8}$$

$$\begin{pmatrix} {}^2R_{11} & {}^2R_{12} & {}^2R_{13} \\ {}^2R_{21} & {}^2R_{22} & {}^2R_{23} \end{pmatrix} \tag{3.9}$$

When the demand of the product is in state 3, two alternatives are open to the decision maker, that is to:

i) No Order

ii) Improve on the method of advertisement

let the corresponding transition probability and reward matrix be given as:

$$\begin{pmatrix} {}^3P_{11} & {}^3P_{12} & {}^3P_{13} \\ {}^3P_{21} & {}^3P_{22} & {}^3P_{23} \end{pmatrix}$$

(3.91)

$$\begin{pmatrix} {}^3R_{11} & {}^3R_{12} & {}^3R_{13} \\ {}^3R_{21} & {}^3R_{22} & {}^3R_{23} \end{pmatrix} \tag{3.92}$$

IV. Application

A business man has found tough competition in regards to the demand of a certain product and would like to use analytical techniques in making decisions for whether to order in large quantity, small quantity or not to order additional units depending on the demands for the product in question. The product undergoes state changes between favorable demand, less favorable and unfavorable states based on the following transition matrices and corresponding reward matrices.

Let the transition probabilities matrix (P_{ij}) and the corresponding reward matrix (R_{ij}) be given as follows:

$$P = P_{ij} = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{pmatrix}; \qquad i, j = 1, 2, 3. \tag{4.1}$$

$$R = R_{ij} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix}; \quad i, j = 1, 2, 3. \tag{4.2}$$

Let D be the decision set and we have two alternative decisions available to the business man. That is, Alternative 1; and Alternative 2; Thus in every state we have k = 1, 2 $\in$ D.

We shall determine the best policies for every n using equation (1.3). Since our interest is to minimize cost and maximize profit, the alternative that yields more earnings constitutes the best policy for the states and time.

The product undergoes state changes base on the following transition probabilities and the corresponding reward matrices in (thousand naira) respectively.

$$P_{ij} = \begin{pmatrix} {}^1p_{11} & {}^1p_{12} & {}^1p_{13} \\ {}^1p_{21} & {}^1p_{22} & {}^1p_{23} \\ {}^1p_{31} & {}^1p_{32} & {}^1p_{33} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.6 & 0.3 \\ 0.3 & 0.5 & 0.2 \end{pmatrix} \tag{4.3}$$

$$i, j = 1, 2, 3 \quad for \quad k = 1$$

$$R_{ij} = \begin{pmatrix} {}^1R_{11} & {}^1R_{12} & {}^1R_{13} \\ {}^1R_{21} & {}^1R_{22} & {}^1R_{23} \\ {}^1R_{31} & {}^1R_{32} & {}^1R_{33} \end{pmatrix} = \begin{pmatrix} 6 & 5 & 4 \\ 4 & 3 & -2 \\ 4 & -6 & 3 \end{pmatrix} \tag{4.4}$$

$$i, j = 1, 2, 3 \quad for \quad k = 1$$

$$P_{ij} = \begin{pmatrix} {}^2p_{11} & {}^2p_{12} & {}^2p_{13} \\ {}^2p_{21} & {}^2p_{22} & {}^2p_{23} \\ {}^2p_{31} & {}^2p_{32} & {}^2p_{33} \end{pmatrix} = \begin{pmatrix} 0.6 & 0.1 & 0.3 \\ 0.5 & 0.3 & 0.2 \\ 0.1 & 0.3 & 0.6 \end{pmatrix} \tag{4.5}$$

$$i, j = 1, 2, 3 \quad for \quad k = 2$$

$$R_{ij} = \begin{pmatrix} {}^2R_{11} & {}^2R_{12} & {}^2R_{13} \\ {}^2R_{21} & {}^2R_{22} & {}^2R_{23} \\ {}^2R_{31} & {}^2R_{32} & {}^2R_{33} \end{pmatrix} = \begin{pmatrix} 6 & 3 & 2 \\ -2 & -10 & 5 \\ -7 & -8 & -1 \end{pmatrix} \tag{4.6}$$

$$i, j = 1, 2, 3 \quad for \quad k = 2$$

We substitute the above values into the optimality equation to obtain our iterations. That is

$${}^oV_i^{(n)} = \max_{k \in D} \left[{}^kQ_i + \sum_{j=1}^{m} {}^kP_{ij} \, {}^oV_j^{(n-1)} \right] \; n = 1, 2, ...; i = 1, 2, ..., m \tag{4.7}$$

V. Results And Discussion

V.

We shall use the values in (4.3) to determine the best policies for every n. where $\sum_{j=1}^{m} P_{ij}\, R_{ij} = Q_i$

we have

$^{1}Q_1 = 5.4$

$^{1}Q_2 = 1.6$

$^{1}Q_3 = -1.2$

$^{2}Q_1 = 4.5$

$^{2}Q_2 = -3$

$^{2}Q_3 = -3.7$

Let $^{o}V_i^{(0)} = 0$ *for* $i = 1,2,3$. *Then for* $n = 1$ in (1.3) *we find* $^{o}V_i^{(1)} = \max_{1,2,3} {}^{k}Q_i$, *hence*

$$d_1^{(1)} = 1,\ d_2^{(1)} = 1 \text{ and } d_3^{(1)} = 1 \tag{5.1}$$

Let $^{o}V_1^{(1)}$, $^{o}V_2^{(1)}$, *and* $^{o}V_3^{(1)}$ be the maximum earnings corresponding to $d_1^{(1)}$, $d_2^{(1)}$ *and* $d_3^{(1)}$ respectively. We have

$$^{o}V_1^{(1)} = 5.4,\ ^{o}V_2^{(1)} = 1.6 \text{ and } ^{o}V_3^{(1)} = -1.2 \tag{5.2}$$

For n = 2, from (3.64) we have

$$^{o}V_i^{(2)} = \max_{1,2,3} \left[{}^{k}Q_i + \sum_{j=1}^{3} {}^{k}P_{ij}\ {}^{o}V_j^{(1)} \right] \tag{5.3}$$

That gives

$i = 1;\ K = 1 \quad {}^{1}V_1^{(2)} = 8.72$

$i = 2;\ K = 1 \quad {}^{1}V_2^{(2)} = 2.74$

$i = 3;\ K = 1 \quad {}^{1}V_3^{(2)} = 0.98$

$i = 1;\ K = 2 \quad {}^{2}V_1^{(2)} = 7.54$

$i = 2;\ K = 2 \quad {}^{2}V_2^{(2)} = -0.06$

$i = 3;\ K = 2 \quad {}^{2}V_3^{(2)} = -3.4$ (5.4)

Clearly,

$$d_1^{(2)} = 1 \text{ with } {}^{1}V_1^{(2)} = 8.72$$

$d_2^{(2)} = 1$ *with* $^{1}V_2^{(2)} = 2.74$

$$d_3^{(2)} = 1 \text{ with } {}^{1}V_3^{(2)} = 0.98 \tag{5.5}$$

Proceeding in this manner, we get

For n = 3:

$$d_1^{(3)} = 1 \text{ with } {}^{1}V_1^{(3)} = 14.696$$

$$d_2^{(3)} = 1 \text{ with } {}^{1}V_2^{(3)} = 5.55$$

$d_3^{(3)} = 1$ *with* $^{1}V_3^{(3)} = 5.162$ (5.6)

For n = 4:

$d_1^{(4)} = 1$ *with* $^{1}V_1^{(4)} = 25.656$

$d_2^{(4)} = 2$ *with* $^{2}V_2^{(4)} = 15.3634$

$d_3^{(4)} = 1$ *with* $^{1}V_3^{(4)} = 13.3812$ (5.7)

For n = 5:

$$d_1^{(5)} = 1 \ \ with \ \ {}^1V_1^{(5)} = 46.79852$$
$$d_2^{(5)} = 2 \ \ with \ \ {}^2V_2^{(5)} = 32.01146$$
$$d_3^{(5)} = 1 \ \ with \ \ {}^1V_3^{(5)} = 31.43594 \qquad (5.8)$$

For n = 6:

$$d_1^{(6)} = 1 \ \ with \ \ {}^1V_1^{(6)} = 87.5668$$
$$d_2^{(6)} = 2 \ \ with \ \ {}^2V_2^{(6)} = 70.451286$$
$$d_3^{(6)} = 1 \ \ with \ \ {}^1V_3^{(6)} = 67.768258 \qquad (5.9)$$

Table 1: The summary result of the optimal policies and rewards

N	$d_1^{(n)}$	$d_2^{(n)}$	$d_3^{(n)}$	${}^oV_1^{(n)}$	${}^oV_2^{(n)}$	${}^oV_3^{(n)}$
1	1	1	1	540	160	-120
2	1	1	1	872	274	980
3	1	2	1	14,696	5,550	5,162
4	1	2	1	25,656	15,3634	13,3812
5	1	2	1	46,7985.2	32,0114.6	31,4359.4
6	1	2	1	87,5666.8	70,4512.86	67,7682.58

The results indicate the best policies for each n. $d_i^{(n)} \ \ where \ n = 1,2,3,4,5,6 \ \ and \ \ i = 1,2,3.$ Thus, we have obtained the best policies for the three states for six months. In addition to the best policies, the corresponding expected rewards are also provided.

For the first month, $d_1^{(1)} = 1 with \ {}^oV_1^{(1)} = 540$ means that the best policy for state 1 is for the business man is to order in large quantity for favorable demand and the corresponding expected reward is five hundred and forty thousand naira.

$d_2^{(1)} = 1; \ with \ {}^oV_2^{(1)} = 160$ Means that the best policy for state 2 is to order small quantity since the demand is less favorable, and the corresponding expected reward is one hundred and sixty thousand naira.

$d_3^{(1)} = 1; \ with \ {}^oV_3^{(1)} = -120$ Means the best policy for state 3 is not to order since the demand is not favorable and the corresponding reward is minus one hundred and twenty thousand naira which is a loss to the business man.

Also for the second month, $d_1^{(2)} = 1; \ with \ {}^oV_1^{(2)} = 872$ means the best policy for state 1 is to keep up advertisement for favorable demand with expected reward of eight hundred and seventy two thousand naira.

$d_2^{(2)} = 1; \ with \ {}^oV_2^{(2)} = 274$ Means the best policy for state 2 is to improve advertisement in order to increase the rate of demand with expected reward of two hundred and seventy-four thousand naira.

$d_3^{(2)} = 1; \ with \ {}^oV_3^{(2)} = -8.2$ Means the best policy for state 3 is not to order since the demand is unfavorable with expected reward of minus eight thousand and two naira.

For the third month, $d_1^{(3)} = 1; \ with \ {}^oV_1^{(3)} = 1,433.6$ means the best policy for state 1 is to order in large quantity with expected reward of one thousand, four hundred and thirty three naira.

$d_2^{(3)} = 2; \ with \ {}^oV_2^{(3)} = 675.8$ Means the best policy for state 2 is to improve on the rate of advertisement and the expected reward is six hundred and seventy-five thousand naira.

$d_3^{(3)} = 2; \ with \ {}^oV_3^{(3)} = 300.2$ Means the best policy for state 3 is not to order with the expected reward of three hundred thousand naira.

The result revealed that for the fourth, fifth and sixth month, the best policies for the states is alternative 1 while for the first state, Alternative 2 for the second state and Alternative 1 for the third state respectively. This is a convergence to stable policy that further iterations beyond is not necessary.

VI. Conclusion

Generally, the importance of decision making in any organization cannot be over emphasized. In this paper, we focus on inventory replenishment based on demands for a particular item as vital aspect of the economy.

This work provides analytic solution to ordering problem of a firm whose aim is to meet up with customer demands, making an optimal choice out of several choices so as to minimize cost and maximize profit. Hence, Markov decision model was analyzed using the value iteration method to achieve optimality in decision making.

From the analysis made, this model can be applied in different areas where decision making is required by altering the elements of the probability and reward matrices depending on the state space of the decision maker and the alternatives.

VII. Recommendation

This research work was done on a three- state model, using the value iteration as method of solution. It recommended that similar work is done by increasing the state space using the policy iteration or linear programmingas a solution method since the value iteration is only applicable when the state space is small.

Reference

[1]. Broekmeullen, R., Van Donselaar, K, Van Woensel T and Fransoo J.C. (2006). Inventory control for perishables in supermarket. Int. J. Prod. Econ. **104**(2), 462–472.

[2]. Cheung, R., and Powell,W. (1996). Models and algorithms for distribution problems with uncertainty demands. Trans. Sci. **30**, 43–59.

[3]. Eynan, A., and Kropp, D. (1998). Periodic review and joint replenishment in stochastic demand environments. IIE Trans. **30**(11).

[4]. Kallen, M.J., van Noortwijk, J.M. (2006), "Optimal periodic inspection of a deterioration process with sequential condition states," International Journal of Pressure Vessels and Piping, vol. 83, no. 4, pp. 249-255.

[5]. Mubiru K.P. and Bernard K.B (2017). The joint location inventory replenishment problem at a supermarket chain under stochastic demand. Journal of Industrial Engineering and Management Science**, 1**,161–178.

[6]. Mubiru K.P, Senfuka Christopher and Ssmpijja Maureen (2019). Modelling cybercafé internet service for revenue optimization under stochastic demand. International Journal of Academic Information Systems Research (IJAISR) Vol**. 3**(5), 32-36

[7]. Puterman, M. (1994). Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley and Sons, New York, NY.

[8]. Piperagkas, G.S., Konstantaras, I, Skouri, K, and Parsopoulos, K.E., (2012) 'Solving the stochastic dynamic lot-sizing problem through nature-inspired heuristics', Computers & Operations Research, **39**, 1555–1565.

[9]. Roychowdhury, S. (2009). An optimal policy for a stochastic inventory model for deteriorating items with time-dependent selling price. Adv.Model. Optim. **11**(3).

[10]. Sloan, T.W., Shanthikumar, J.G.,(2002) "Using in-line equipment condition and yield information for maintenance scheduling and dispatching in semiconductor wafer fabs," IIE Transactions, vol. 34, no. 2, pp. 191-209.

[11]. Saranga, H., Knezevic, J., (2001) "Reliability prediction for condition-based maintained systems," Reliability Engineering and System Safety, vol. 71, no. 2, pp. 219-224.

[12]. Taylor, H. M., &Karlin, S. (1998). An introduction to stochastic modeling. Boston: Academic Press

[13]. Yin, K.K., Liu, H., and Johnson, E.N., (2002) 'Markovian inventory policy with application to the paper industry', Computers and Chemical Engineering, **26**, 1399-1413.

[14]. Yu, C.-S. and Li, H.-L. (2000). A robust optimization model for stochastic logistic problems. International Journal of Production Economics**. 64** (1-3), 385-397.

[15]. Yin, G., Zhang, Q., Yang, H., & Yin, K. (2001). Discrete-time dynamic systems arising from singularly perturbed Markov chains. Nonlinear analysis: theory, methods & applications, 47 (7),4763–4774.

[16]. Yin, G. G., & Zhang, Q. (1998). Continuous Markovchains and applications, a singular perturbation approach. New York: Springer.

[17]. Yin, G. G., & Zhang, Q. (1997). Mathematics of stochastic manufacturing systems. Providence: American Mathematical Society.

[18]. Yin, K., Yin, G. & Zhang, Q. (1995). Approximating the optimal threshold levels under robustness cost criteria for stochastic manufacturing systems, Proceeding IFAC Conference of Youth AutomationYAC '95, 450–454

[19]. Zhao, Q. H., Chen, S., Leung, S. C. H. and Lai, K. K. (2010). Integration of inventory and transportation decisions in a logistics system. Transportation Research Part E: Logistics and Transportation Review. **46** (6), 913-925.

[20]. Zheng,Y. (1992). On properties of stochastic inventory systems. Manag. Sci. 38, 87–103..Oper. Res. **203**, 55–80.

Aisha Sheikh Hassan, et. al. "An Optimal Ordering Policy with Markov Decision Process." *IOSR Journal of Mathematics (IOSR-JM)*, 16(3), (2020): pp. 08-17.