NONLINEAR AND DYNAMIC PROGRAMMING MODELS FOR AN INVENTORY PROBLEM IN A PARTIALLY OBSERVABLE ENVIRONMENT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ALP DARENDELİLER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

JANUARY 2016

Approval of the thesis:

NONLINEAR AND DYNAMIC PROGRAMMING MODELS
FOR AN INVENTORY PROBLEM
IN A PARTIALLY OBSERVABLE ENVIRONMENT

submitted by **ALP DARENDELİLER** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Murat Köksalan
Head of Department, **Industrial Engineering** _______________

Prof. Dr. Yasemin Serin
Supervisor, **Industrial Engineering Dept., METU** _______________

Examining Committee Members:

Assoc. Prof. Dr. Seçil Savaşaneril
Industrial Engineering Dept., METU _______________

Prof. Dr. Yasemin Serin
Industrial Engineering Dept., METU _______________

Assoc. Prof. Dr. İsmail S. Bakal
Industrial Engineering Dept., METU _______________

Assist. Prof. Dr. Sakine Batun
Industrial Engineering Dept., METU _______________

Assist. Prof. Dr. Banu Yüksel Özkaya
Industrial Engineering Dept., Hacettepe University _______________

Date: 21 January 2016

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Alp Darendeliler

Signature :

ABSTRACT

NONLINEAR AND DYNAMIC PROGRAMMING MODELS
FOR AN INVENTORY PROBLEM
IN A PARTIALLY OBSERVABLE ENVIRONMENT

Darendeliler, Alp
M.S., Department of Industrial Engineering
Supervisor: Prof. Dr. Yasemin Serin

January 2016, 46 Pages

In this study, a single-item periodic-review inventory system is considered in a partially observable environment with finite capacity, random yield and Markov modulated demand and supply processes for finite-horizon. The exact state of the real process, which determines the distribution of the demand and supply, is unobservable so the decisions must be made according to the limited observations called observed process. Partially Observable Markov Decision Process is used to model this problem. As an alternative to the dynamic programming model, a nonlinear programming model is developed to find optimal policies. The optimal policies of the nonlinear program is more practical to obtain and use compared to the dynamic programming model. Computational study is performed for the three data sets in order to compare the results of the two models. The results show that the optimal policies of the two models are the same.

Keywords: Markov-modulated demand, Markov-modulated supply, Partially Observable Markov Decision Process

ÖZ

KISMİ GÖZLEMLENEBİLEN ORTAMDAKİ ENVANTER PROBLEMİ İÇİN DİNAMİK VE DOĞRUSAL OLMAYAN MODELLER

Darendeliler, Alp
Yüksek Lisans, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Yasemin Serin

Ocak 2016, 46 Sayfa

Bu çalışmada, kısıtlı olarak gözlenebilen ortamda, kapasite sınırlı, tek parça, periyodik olarak incelenen ve Markov-modüle talep ve arz dağılımlı olan sonlu zamanlı envanter sistemi incelenmiştir. Talep ve arz dağılımlarını belirleyen sürecin durumu tamamen gözlenememektedir. Bu nedenle, kararlar gözlenen sürece göre verilmektedir. Problemi modellemek için Kısmi Gözlemlenebilen Markov Karar Süreci kullanılmıştır. Optimum sonuçları bulmak için dinamik programlama modeline alternatif olarak doğrusal olmayan bir model geliştirilmiştir. Modelden elde edilen optimum kararların bulunması ve uygulanması dinamik programlamaya göre daha pratiktir. Belirtilen iki modelden çıkan sonuçları karşılaştırmak için üç veri seti üzerinden sayısal çalışma yapılmıştır. Bu çalışmalardan çıkan sonuçlar birbirinin aynısıdır.

Anahtar Kelimeler: Markov-modüle talep, Markov-modüle arz, Kısmi Gözlemlenebilen Markov Karar Süreci

ACKNOWLEDGMENTS

The author wishes to express his deepest gratitude to his supervisor Prof. Dr. Yasemin Serin for her guidance, advice, encouragements and insight throughout the research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

$Z_{\boldsymbol{t}}$	Real process at time t
$Y_{\boldsymbol{t}}$	Observed process at time t
$Q_t(a, b)$	Probability that next state of the real process is a given that the current state is b
$R_t(a, i)$	Probability of observing state i when the actual state is a
π_t^a	Probability that actual state is a given that all of the observations up to time t
D_t	Random demand at time t
U_t	Random yield at time t
x_t	On-hand inventory at the beginning of time t
y_t	Order-up-to level at time t
K	Capacity of the producer
c_i	Purchasing cost if the observed state is i
h_i	Holding cost if the observed state is i
p_i	Shortage cost if the observed state is i
γ	Discount factor
$J_t(i, a, x, y)$	Total discounted expected cost at time t when the actual state is a, observed state is i, inventory level is x and order-up-to level is y
$v_t(\pi, i, x)$	Optimal cost function at time t for a given $\pi, i\ and\ x$
$y_t(\pi, i, x)$	Optimal order-up-to level at time t for a given $\pi, i\ and\ x$
$s_t^{\pi,i}$	Threshold level at time t for a given π and i; below this level it is optimal to order the capacity
$S_t^{\pi,i}$	Threshold level at time t for a given π and i; above this level it is optimal to not to order
$V_{(a,i,x,t)}$	Total discounted expected cost at time t when the actual state is a, observed state is i, inventory level is x (Used in nonlinear

program)

$\alpha_{(a,i,x,y,t)}$	Probability of taking action y when the actual state is a, observed state is i, inventory level is x and order-up-to level is y (Used in nonlinear program)
$c_{(a,i,x,y)}$	One-period expected cost when the actual state is a, observed state is i, inventory level is x and order-up-to level is y (Used in nonlinear program)

CHAPTER 1

PROBLEM DEFINITON AND LITERATURE REVIEW

1.1 PROBLEM DEFINITION

In this study, a single-item periodic-review inventory system is considered in a randomly changing environment, where demand, supply and cost parameters are affected by it. Outside environment is influenced by many factors such as economic, political, social and financial conditions. These factors are modeled by a discrete time discrete state Markov chain called real process that represents the state of the outside world. Decision maker cannot identify all of the factors that determine the state of the real process so he has incomplete information about it. He must make ordering decisions based on his limited observations. These observations are defined as observed process which does not necessarily satisfy Markov property. For instance, a production company may not have adequate information about the economy to estimate the distribution of demand of its product. Instead, it could obtain partial information about the state of the economy. Based on this, the company can determine how much to produce in order to minimize its expected total inventory cost.

It is assumed that demand is random and it is modulated by the state of the real process. In different states of the unobserved (real) process, the distribution of demand varies. For instance, if the state of the economy is good, demand follows a distribution with a high mean. In contrast, in case of economic decline, demand has distribution with a lower mean.

Besides, inventory manager may face random yield in some cases and a random proportion of the order can be received. The factors that cause yield randomness are breakdowns, repairs, maintenance, learning and introduction of new technologies.

The aim of this study is to find practical policies for finite-horizon inventory system which is subject to random yield, fixed capacity and uncertain demand in a partially observed environment. Besides, the impact of the randomly changing environment on demand, supply and cost parameters and also on optimal policies are analyzed. The outside world is modeled by a Markov chain called real process. It cannot be fully observed. Demand and supply are modulated by this process. Decision maker has imperfect information and the observations he made are defined by an observed process which is probabilistically related to the real process. Inventory costs are also affected by the external factors. Partially Observable Markov Decision Process (POMDP) is used to find and analyze the optimal policies for the problem. POMDP's are generally used for modeling sequential decision processes. It consists of set of states, set of actions, transition probabilities, costs, set of observations and discount factor. Internal process cannot be fully observed. Instead, decision maker has imperfect observations called observable process and it is probabilistically related to the unobserved process.

1.2 LITERATURE REVIEW

Demand has been modeled by Markov modulated process by many authors so as to consider the effect of the fluctuating environment on demand. Because outside environment influenced by many factors such as economic conditions, price competition and product differentiation, modeling demand as stationary is not sufficient to reflect demand uncertainty. Therefore, outside environment is represented by *state-of-the-world* of a Markov process in order to consider the effect of outside environment on demand.

Karlin and Fabens (1959) consider an inventory model in which the distribution of the demand is determined by the state of the Markov chain. They prove that different (s, S) type inventory policies are optimal for each state of the Markov chain. Iglehart and Karlin (1962) study single-item periodic-review inventory model with non-stationary stochastic demands for finite-horizon and analyze the structure of the optimal policies. They model demand as a Markovian process. Distribution of demand is different for each state of the Markov Chain. They found out that base-stock policy is optimal where critical numbers are determined by the state of the demand process under convexity assumptions.

Sethi and Cheng (1993) extend the Karlin and Fabens (1959) study. They model outside environment by Markov chain as in the previous studies. The inventory/backlog costs are state-dependent and convex. They prove that state-dependent (s, S) policies are optimal for finite horizon inventory model.

Song and Zipkin (1993) also study Markov modulated demand process. They model "outside world" as a continuous-time, discrete state space Markov process. It is assumed that demand distribution, whose parameter depends on "outside world", is Poisson. They allow countably-infinite state space, stochastic lead times, fixed and linear order costs as an extension to the Iglehart and Karlin (1962). The world dependent base-stock policy is optimal where order cost is linear in quantity. In the case of fixed ordering cost, world-dependent (s, S) policy is optimal where s and S are functions of the state of the Markov process representing outside environment.

Another issue in inventory control is supply uncertainty. Supply uncertainties are generally considered in two groups: random yield, random capacity in literature. Henig and Gerchak (1990) studied the effect of the yield randomness on the optimal policies for single-item periodic-review inventory model for single-period, finite-horizon and infinite horizon. They assumed that the environment is stationary, demand is uncertain and yields are stochastically proportional to input level. For the single period, order point is same as the constant yield case under general conditions. The optimal policy is called nonorder-up-to policy for finite-horizon. They also show that the infinite-horizon order point is higher than when there is no proportional yield.

Federgruen and Zipkin (1986) studied single-item, periodic-review inventory model with uncertain demands where the order capacity is fixed. They assumed that demand is stationary, cost functions are convex and distribution of the demand is continuous. They found out that modified base-stock policy is optimal for both finite and infinite horizons. In this policy structure, it is optimal to order up to the critical number or order as close to it as possible when the initial stock is below that number. If the amount of initial inventory exceeds base-stock level, no order should be given.

Uncertain capacity in production planning is studied by Ciarallo et al. (1994). They analyze the effect of random capacity on production planning decisions for single-period, finite-horizon and infinite-horizon in a stationary environment. Demand is uncertain and the costs are linear. Optimal policy, which is identical to the classic newsboy problem, is not influenced by the random capacity for the single-period problem. The cost function is unimodal and nonconvex. Base-stock policy is shown to be optimal for finite-horizon and infinite-horizon problems.

Wang and Gerchak (1996) study both the variable capacity and random yield for the production planning problem where demand is stationary and uncertain. They analyze optimal policies for finite-horizon and infinite-horizon. It is showed that the objective function is quasi-convex. The optimal policy that minimizes total expected discounted cost is characterized by a single critical point for finite-horizon problem. However, the structure of the optimal solution is not base-stock policy. It differs from the work of Ciarallo et.al (1994) where variable capacity was considered as the only supply uncertainty.

Later, studies have considered the effect of fluctuating environment on supply. Özekici and Parlar (1999) consider infinite-horizon periodic-review, single-item inventory model with unreliable suppliers. Demand, supply and cost parameters are driven by the randomly changing environment which is modeled by time-homogeneous Markov chain. Distribution of demand depends on this process. Yield randomness is modeled different from the previous random yield supply models. Amount of order received is either full with probability p_i or nothing is delivered with probability $(1 - p_i)$. The probabilities change according to the state of the environment. When the inventory ordering cost is linear with ordered quantity, state-dependent base-stock policy is optimal. If there is also a fixed ordering cost, then the optimal solution is state-dependent (s, S) policy. Erdem and Özekici (2002) extends Özekici and Parlar (1999). They analyzed similar problem for single-period, finite-horizon and infinite-horizon. However, they modeled supply uncertainty by random yield and supplier is assumed to be always available. In all cases, the optimal policy is state-dependent base-stock policy. Order-up-to levels differs according to the state of the environment.

Gallego and Hu (2004) analyze single-item, periodic review inventory problem with random yield and finite capacity. Distribution of the demand and supply depend on two time-homogeneous Markov chains. Decision maker chooses actions according to the state of the demand and yield. Gallego and Hu defines optimal policy as "modified state dependent inflated base-stock policy". In this policy structure, there are two critical values $(s^{(i,a)}, S^{(i,a)})$, which depend on the current state of the supply and demand processes, determine the optimal policy. If the inventory level is below $s^{(i,a)}$, then it's optimal to order the capacity. If it is above $S^{(i,a)}$, then nothing should be ordered. For the case where the inventory level is between these two critical values, optimal order-up-to level is greater and equal to the $S^{(i,a)}$. Besides, it is decreasing in the amount of inventory level.

Finding optimal policies for unobservable Markov chains with finite states for finite horizon is first introduced by Smallwood and Sondik (1971). They redefine the process as an observable chain over a continuous state space. The optimal policies were difficult to find and essentially not deterministic over the original state space. Treharne and Sox (2002) analyzed the finite-horizon inventory control problem where the demand is nonstationary and partially observed. They do not consider supply uncertainty in this work. It is assumed that distribution of the demand is determined by the state of the Markov chain called core process. This process has finite number of states and each of them generate different demand distribution. Decision maker cannot exactly determine which of the distributions generate the demand in any period. He can observe only the actual demand of the previous periods. State-dependent base-stock policy is shown to be optimal.

Arifoğlu and Özekici (2010) use POMDP to model the uncertain demand, finite capacity and random yield for single-period, multiple-period and infinite-period problems. Demand and supply processes are modulated by the Markov Chain called real (unobserved) process. Decision maker cannot directly observe this process. They show that state-dependent modified inflated base-stock policy is optimal. Another study of Arifoğlu and Özekici (2011) also use POMDP for the modeling inventory problem. They consider supply uncertainty in terms of random capacity of production and random availability of transportation. They analyze optimal policies for single-period, multiple-period and infinite-period problems. In these POMDP based studies, since the state variable has current probability distribution of the states, the optimal policy is nondeterministic with respect to the original states.

Serin and Avşar (1997) use nonlinear programming model to find optimal policies for the Markov Decision Process with restricted observations for finite horizon. In their study, internal process cannot be observed but the states are grouped so that the group that a state belong to is observable. Decisions are made according to the groups of the states so as to minimize expected total discounted cost. This model can be used as an approximation to the POMDP models. The optimal policy is guaranteed to be deterministic over the original state space.

In the present study, two models in Arifoğlu and Özekici (2010) and Serin and Avşar (1997) for a finite horizon inventory problem are compared in terms of their solutions. The performance of the second as an approximation to the first is analyzed.

CHAPTER 2

MODELS

2.1 PROBLEM

A single-item periodic-review multi-period inventory system is considered in a partially observable environment with finite capacity, random yield and Markov modulated demand and supply processes as in Arifoğlu and Özekici (2010). Outside environment is modeled by discrete-time discrete-state Markov chain called real process that represents the state of the outside world whose state generates the demand and supply distributions. The state of the real process is not exactly known so decision maker has imperfect information. The ordering decisions must be made based on the limited observations. These observations are defined as observed process which need not to satisfy Markov property. The purpose is to find practical policies for finite-horizon inventory system under these conditions. Arifoğlu and Özekici (2010) solved this problem using dynamic programming model with a continuous-state space.

In the present work, the nonlinear programming model in Serin and Avşar (1997) which is used for solving POMDP for finite-horizon, is adapted to model and solve the inventory problem posed in Arifoğlu and Özekici (2010). Since the policies that are obtained from nonlinear programming model are deterministic with respect to the original states of the problem, they are more practical to find and apply compared to the dynamic programming model. Next, these two models are presented.

2.2 DYNAMIC PROGRAMMING MODEL

A T-period dynamic programming model is constructed for a single-item periodic-review inventory system under partially observable environment in Özekici and Arifoğlu (2010). Real state Z_t, is a discrete-time and time-dependent Markov chain with state space $F = \{a, b, \ldots\}$ and transition matrix $Q_t(a, b)$. Environment is represented by Z_t which directly affects demand and supply distributions. The states of Z_t cannot be observed since it depends on many factors such as economic, financial, political and other factors. Therefore, true state of the environment is unobservable. However, there is an observed process called Y_t with state space $E = \{i, j, ..\}$ and it is probabilistically related to the real state. Sondik and Smallwood (1971) calls Z_t as "Internal Markov Process" and Y_t as "Observable Output". The conditional probability of Y_t given Z_t is

$$P\{Y_t = i | Z_t = a\} = R_t(a, i). \tag{2.1}$$

$\boldsymbol{Y_t} = (Y_o, Y_1, \ldots, Y_t)$, represents all of the observations up to time t; π_t^a indicates the probability of state a of the unobserved real state at time t given all of the observations until time t; π_t^a is

$$\pi_t^a = P\{Z_t = a | \boldsymbol{Y_t}\} \qquad a \in F \tag{2.2}$$

and

$\pi_t = [\pi_t^a, \pi_t^b, \ldots]$ is the conditional distribution of Z_t given $\boldsymbol{Y_t}$ satisfying

$$\sum_{a \in F} \pi_t^a = 1 \quad \text{where } \pi_t^a \geq 0 \quad for\ t = 1, \ldots, T \tag{2.3}$$

At time t, information vector π, is updated according to the observed process Y_{t+1} by the equation

$$\pi_{t+1}^{b} = P\{Z_{t+1} = b | Y_{t+1} = j\} = \frac{\sum_{a\in F} \pi_t^a Q_t(a,b) R_{t+1}(b,j)}{\sum_{a,c\in F} \pi_t^a Q_t(a,c)\, R_{t+1}(c,j)} \tag{2.4}$$

and $\{\pi_t\,;\ t \geq 0\}$ is a fully observable continuous state Markov process.

Demand process, denoted by $D_t = \{D_t;\ t \geq 1\}$, depends on unobserved state Z_t. The conditional cumulative distribution function of the demand given the real state is

$$M_a(z) = P\{D_{t+1} \leq z | Z_t = a\} \text{ for } a \in F \text{ and } z \geq 0. \tag{2.5}$$

Let x_t denotes the inventory level and y_t denotes order-up-to level at time t. At each time period, inventory manager places an order of $y_t - x_t$ to the supplier. Here, the order up-to level y_t is the action taken at time t, of the dynamic programming model. Inventory manager cannot receive the order completely because of the factors such as defective production, transportation problems. $U_t \in [0,1]$ is the proportion of the order that is received at time t. The cumulative distribution of U_t for a given real state a at time t is

$$F_a(u) = P\{U_{t+1} \leq u | Z_t = a\}. \tag{2.6}$$

The conditional expectation of the random yield is

$$\mu_a = E[U_{t+1} | Z_t = a]\ , \tag{2.7}$$

Let K be the capacity of the supplier that is the bound of the order. Then, the amount of random order received in period t is

$$S_t(y_t - x_t) = U_{t+1} \min\{y_t - x_t, K\}. \tag{2.8}$$

$\mu_{\boldsymbol{\pi}}$ is assumed to be greater than zero which means that there must be a state a such that $\mu^a > 0$, with the probability $P\{U_{t+1} = 0 | Z_t = a\} < 1$. In this model, cost parameters are not fixed. All the costs depend on the state of the observable process Y_t since inventory manager can only see this process. The purchasing, holding and shortage cost are defined as c_i, h_i and p_i respectively given that the observed process is i. Holding and shortage costs are incurred at the end of the each period. Purchasing cost is incurred at the beginning of each period. It is computed with respect to quantity received not to the quantity ordered. Moreover, it is assumed that $p_i > c_i > 0$ and $h_i > 0$ for all i and all of them are finite. The discount factor γ is in interval $(0,1)$ which is used to compute multi-period total cost.

Lead time is assumed to be zero for the computations. In addition, all of the unsatisfied demand is backlogged. To compute π_{t+1}, Q_t, Y_{t+1}, R_{t+1}, π_t are sufficient as the Equation (4) illustrates. Hence, the problem can be modeled as a completely observable discrete-time continuous state Markov Decision Process with states (π_t, x_t, y_t).

Assuming that the inventory level at the beginning of period is x, the initial distribution is π, the observed state is i, and the actual state is a, the one-period costs are computed by the equations:

$$L(i,a,y) = h_i \int_0^y (y-z) dM_a(z) + p_i \int_y^\infty (z-y) dM_a(z) \tag{2.9}$$

$$G_0(i,a,y) = c_i y + L(i,a,y) \tag{2.10}$$

$$J_0(i,a,x,y) = \int_0^1 G_0\big(i,a,x+u(y-x)\big) dF_a(u) - c_i x \tag{2.11}$$

$$H_0(\pi,i,x,y) = \sum_{a \in F} \pi^a J_0(i,a,x,y). \tag{2.12}$$

One-period total holding and shortage costs are computed by Equation (9). Equation (11) computes the one-period expected total costs for states (i,a,x) and order-up-to level y by considering proportional yield denoted by u. Because actual state a cannot be exactly observed, the one-period total expected cost according to the information vector π is calculated by Equation (12).

Single period minimum cost function $v_0(\pi, i, x)$ is

$$v_0(\pi, i, x) = \min_{x \le y \le x+K} H_0(\pi, i, x, y)\ . \tag{2.13}$$

Arifoğlu and Özekici (2010) prove that optimal ordering policy for the single-period model is a state-dependent modified inflated base-stock policy

$$y_0(\pi, i, x) = \begin{cases} x + K, & x < s_0^{\pi,i}, \\ y_0^{\pi,i}(x), & s_0^{\pi,i} \le x < S_0^{\pi,i}, \\ x & x \ge S_0^{\pi,i}. \end{cases} \tag{2.14}$$

where single-period variables $y_0^{\pi,i}(x)$, $S_0^{\pi,i}$ and $s_0^{\pi,i}$ are unique y values that are obtained from the solution of

$$\sum_{a \in F} \pi_0^a \int_0^1 (c_i + L'\,(i, a, x + u(y - x)))dF_a(u) = 0, \tag{2.15}$$

$$\sum_{a \in F} \pi_0^a\ \mu_a\big(c_i + L'(i, a, y)\big) = 0, \tag{2.16}$$

$$\sum_{a \in F} \pi_0^a \int_0^1 (c_i + L'\,(i, a, x + uK))dF_a(u) = 0. \tag{2.17}$$

for all π_0 and i respectively. In this policy structure $S_0^{\pi,i}$ depends on the mean of the proportional yield and optimal order-up-to level is nonincreasing for all $\pi, i\ and\ x\ \in [s_0^{\pi,i}, S_0^{\pi,i})$. In case of $U = 1$ with probability 1, all of order can be received and optimal policy is independent of the $x \in [s_0^{\pi,i}, S_0^{\pi,i})$. This type of policy is called state-dependent modified base-stock policy in which it is optimal to order to the amount $S_0^{\pi,i}$ when $x \in [s_0^{\pi,i}, S_0^{\pi,i})$. Otherwise, same policy is applied as in previous case.

For the multi-period problem, dynamic programming equation involves the single period costs plus expected discounted cost of the next period. Assuming that the inventory level at the beginning of the horizon is x, the observed state is i, the real state is a, and the information vector is π at time t, then multi-period costs are computed by the equations:

$$G_t(\pi, i, a, y) = c_i y + L(i, a, y) + \gamma \sum_{j \in F} P\{Y_{t+1} = j | X_t = a\} E_D^a [v_{t+1}(\pi_{t+1}, j, y - D)] \tag{2.18}$$

$$J_t(i, a, x, y) = \int_0^1 G_t\big(\pi, i, a, x + u(y - x)\big) dF_a(u) - c_i x \tag{2.19}$$

$$H_t(\pi, i, x, y) = \sum_{a \in F} \pi^a J_t(\pi, i, a, x, y). \tag{2.20}$$

At time t when there is $(T - t)$ periods remaining, the one-period holding and shortage costs plus the expected total discounted cost of the remaining periods is calculated by equation (18). Equation (19) computes the total discounted expected cost when the state is (i, a, x) and order-up-to level is y. Since actual state a cannot be observed, the multiple-period total expected cost according to the information vector π, is calculated by Equation (12). E_D^a is the expectation with respect to the random demand D for the unobserved state a.

Multiple period optimal cost function satisfies:

$$v_t(\pi, i, x) = \min_{x \le y \le x + K} H_t(\pi, i, x, y). \tag{2.21}$$

It is strictly convex which is proved by Özekici and Arifoğlu (2010). Özekici and Arifoğlu (2010) prove that the optimal ordering policy for the T-period model is a state-dependent modified inflated base-stock policy

$$y_t(\pi, i, x) = \begin{cases} x + K, & x < s_t^{\pi,i}, \\ y_t^{\pi,i}(x), & s_t^{\pi,i} \leq x < S_t^{\pi,i}, \\ x & x \geq S_t^{\pi,i}. \end{cases} \tag{2.22}$$

where $y_t^{\pi,i}(x)$, $S_t^{\pi,i}$, $s_t^{\pi,i}$ are the y values which uniquely satisfy

$$\sum_{a \in F} \pi^a \int_0^1 (\, G_t'(\pi, i, a, x + u(y - x)))dF_a(u) = 0, \tag{2.23}$$

$$\sum_{a \in F} \pi^a \, \mu_a G_t'(\pi, i, a, y) = 0, \tag{2.24}$$

$$\sum_{a \in F} \pi^a \int_0^1 (uG_t'\,(\pi, i, a, y + uK))dF_a(u) = 0, \tag{2.25}$$

for all π and i respectively. The optimal policy structures are the same as in the single-period model for the case of no proportional yield and random yield.

Note that critical values are computed by the dynamic programming algorithm not from the solution of these equations.

2.3 A NONLINEAR PROGRAMMING MODEL FOR FINITE HORIZON PARTIALLY OBSERVABLE MARKOV DECISION PROCESS

A fully observable Markov Decision Process consists of a set of states F, a set of actions A, the expected cost of taking action y in state a is c_{ay}. Z_t denotes the state of the system which is fully observed and A_t is the action taken at time t. The probabilities of one-step transition probability matrix $P(y)$ of the MDP under action a is stated as:

$$p_{ab}(y) = P\{Z_{t+1} = b \mid Z_t = a, A_t = y\} \qquad for\ all\ a\ , b \in \text{F, y} \in A. \tag{2.26}$$

There are T time periods in which decision maker chooses an action at each decision epoch according to the past states and observations $(Z_1, A_1, \dots, Z_T, A_T)$. The cost incurred at every each period is $C(Z_t, A_t)$. Its expected value given that current state and action is

$$c_{ay} = E[C(Z_t, A_t) \mid Z_t = a, A_t = y] \quad \textit{for all } a \in F, y \in A, t = 1, \dots, T \tag{2.27}$$

Let V_{at} be total discounted cost at time t when there is $(T - t)$ time periods remaining and the state is a. The expected T period total discounted cost, starting in an initial state a is

$$V_{a0} = c_{ay} + \gamma \sum_{b=1}^{N} p_{ab}(y)\, V_{b1}. \qquad for\ all\ \ y \in A \tag{2.28}$$

where γ is the discount factor which is between 0 and 1. The total expected discounted cost is computed by the equation:

$$L = \sum_{a \in F} p_0 V_{ao} \tag{2.29}$$

where the initial distribution $p_a = P\{Z_0 = a\}\ for\ a \in F,$ is the probability that the system is in each state at the beginning of the process, is assumed to be known. The sum of all p_a's are equal to 1.

For this finite horizon Markov Decision Process problem, the objective is finding the optimal policy which minimizes L. There are three main approaches to find the optimal solution: policy-iteration algorithm, value-iteration algorithm and linear programming are used to solve the problem. In this study, linear programming is considered. The optimal total discounted cost is

$$V_{at}{}^{*} = \min_{y \in A}\{\, c_{ay} + \gamma \sum_{t=1}^{T} p_{ab}(y) V_{b(t+1)}{}^{*}\} \quad for\ all\ a \in F\ and\ t = 1 \dots, T \tag{2.30}$$

The linear program for the finding optimal policy of a fully observable MDP over finite horizon is expressed as in Serin and Avşar (1997)

$$maximize \quad z = \sum_{a=1}^{N} p_0 V_{a0}$$

$$subject\ to$$

$$V_{at} \leq c_{ay} + \gamma \sum_{b \in F} p_{ab}(y) V_{b(t+1)} \qquad for\ all\ a \in F,\ y \in A, t = 1, \dots, T-1 \tag{2.31}$$

$$V_{aT} = 0 \qquad\qquad for\ all\ a \in F. \tag{2.32}$$

In case of partial observability, actual state Z_t cannot be observed. Instead, there is an observable process Y_t that gives partial information about the real state. Unlike the POMDP methodology adopted in the above dynamic programming model, the decisions may be based on only the current observation in order to obtain more practical policies to apply. This is a suboptimal but easy to use approximation to the POMDP solution. The following nonlinear program is adapted from the study of Serin and Avşar (1997) for this purpose.

A new Markov decision process with state (Z_t, Y_t) where Z_t represents the unobservable state of the original process taking values in $F = \{a, b, ...,\}$ and Y_t is the observable process taking values in $E = \{i, j, ...\}$ and Y_t need not to be Markovian. Let the action space be $A = \{1,2, ..., M\}$. In that case a policy vector $\boldsymbol{\alpha}$ is defined with the following the probabilities: the probability of taking action a when the state is $(Z_t = a, Y_t = i)$ is

$$\alpha_{(a,i,y,t)} = P\{A_t = y | Z_t = a, Y_t = i\}. \qquad (2.33)$$

In this new Markov decision process, the decision maker must take an action with the same probability whenever he observes a state (Z_t, Y_t) with the same observation value Y_t:

$$\alpha_{(a,i,y,t)} = \alpha_{(b,i,y,t)}. \qquad for\ all\ a, b \in F, i \in E, y \in A, t = 1, ..., T \qquad (2.34)$$

These implicit constraints are called observability constraints since it limits actions according to the observations.

$$F = \{\alpha \in R^{|F||E|T} : \sum_{y=1}^{M} \alpha_{(a,i,y,t)} = 1 \qquad for\ all\ \ i \in E\ t = 1, ..., T\ \ and \qquad (2.35)$$

$$\alpha_{(a,i,y,t)} \geq 0 \qquad for\ all\ a \in F, j \in E, y \in A, t = 1, .., T \qquad (2.36)$$

The probability of transition to the state $(Z_t = b, Y_t = j)$ from the state $(Z_{t-1} = a, Y_{t-1} = i)$ under policy $\boldsymbol{\alpha}$ is

$$p_{(a,i),(b,j)}(\boldsymbol{\alpha}, t) = \sum_{y \in A} \alpha_{(a,i,y,t)}\, p_{(a,i),(b,j)}(y) \qquad (2.37)$$
$$for\ all\ (a,b) \in F, (i,j) \in E\ and\ t = 1, \dots, T.$$

The cost incurred at the beginning of each period depends on the state of the system and the action taken. That is

$$c_{(a,i,t)}(\boldsymbol{\alpha}) = \sum_{y \in A} \alpha_{(a,i,y,t)} c_{ay}. \qquad for\ all\ a \in F, i \in E, t = 1, \dots, T \qquad (2.38)$$

where c_{ay} is computed by the equation (24).

Total expected cost from the beginning of period t to the last period T starting from the state (a, i) is denoted by V_{ait}. It is expressed as

$$V_{ait} = c_{(a,i,t)}(\boldsymbol{\alpha}) + \gamma \sum_{j \in E} \sum_{b \in S} p_{(a,i),(b,j)}(\boldsymbol{\alpha}, \boldsymbol{t})\ V_{bj(t+1)} \qquad (2.39)$$

The model program for the new Markov Decision Process minimizing expected cost over a T-period horizon is

$$minimize \qquad \sum_{a \in S} \sum_{i \in E} p_0\, V_{ai0}$$
$$subject\ to \qquad (2.40)$$
$$V_{ait} = c_{(a,i)t}(\boldsymbol{\alpha}) + \gamma \sum_{j \in E} \sum_{b \in S} p_{(a,i),(b,j),}(\boldsymbol{\alpha}, \boldsymbol{t})\ V_{bj(t+1)}$$
$$for\ all\ a \in F, i \in E, t = 0, \dots, T-1$$

.

$$\sum_{y \in A} \alpha_{(a,i,y,t)} = 1 \qquad for\ all\ a \in F, i \in E, t = 0, \dots, T-1 \qquad (2.41)$$

$$\alpha_{(a,i,y,t)} = \alpha_{(b,i,y,t)} \qquad for\ all\ a, b \in F, i \in E, y \in A, t = 1, \dots, T \qquad (2.42)$$

$$\alpha_{(a,i,y,t)} \geq 0 \qquad for\ all\ a \in F, i \in E, y \in A, t = 0, \dots, T-1 \qquad (2.43)$$

$$V_{aiT} = 0 \qquad for\ all\ a \in F, i \in E. \qquad (2.44)$$

Note that both $\boldsymbol{\alpha}$ and V's are unknown and the above model is a nonlinear program.

Equation (41) provides that the sum of the probabilities for taking actions $y \in A$ for each of the state (a, i, t) is equal to 1. The same probability of taking action y for the same observed state i is ensured by equation (42). It is stated that the discounted cost variable V_{aiT} at time T, when there is no period remaining, is equal to 0 in equation (42).

Serin and Avşar (1997) proved that the optimal policy is deterministic. This means that $\alpha_{(a,i,y,t)} \in \{0,1\}$. Hence this nonlinear program can be solved with binary variables.

2.4 PARAMETERS OF THE NONLINEAR PROGRAM FOR THE PARTIALLY OBSERVABLE MULTIPLE-PERIOD INVENTORY PROBLEM

Multiple-period inventory problem in a partially observable environment can be solved by the previous nonlinear program. The notations used in this part are the same as the Dynamic Programming Model. Z_t is the real process and Y_t is the observed process. $R_t(a, i)$ denotes the same probability as in (1) and $Q_t(a, b)$ represents transition probabilities of Z_t $for\ all\ a, b \in F$. States of the Markov Decision Process are denoted by (Z_t, Y_t, x_t), where x_t is the inventory level at the beginning of period t. It can take negative values because unsatisfied demand is backlogged. $x \in X$ is the discrete set of inventory levels Actions are the amount of the inventory that is to be ordered. The order up to level for each period t is denoted by y_t which is also an action for this MDP. At time t, actions space must be in the interval ($x_t \leq y_t \leq x_t + K$). It cannot exceed on-hand inventory plus capacity of the order.

The continuous demand is discretized in order to use in the nonlinear model. The demand distribution depends on the state of the unobservable process Z_t. So, let $P(D^a = d)$ be the discrete probability mass function of the demand if the unobservable state is a, and d takes values according to the discretization. Discrete approximation is also made for the random yield U. Let $\mathrm{P}(U^a = u)$ be the discrete probability mass function of the demand if the unobservable state is $a, a \in F$, and d takes values according to the discretization

Transition probability from the state $(Z_t = a, Y_t = i, x_t)$, to the $(Z_{t+1} = b, Y_{t+1} = j, x_{t+1})$ under action y can now be computed as:

$$p_{(a,i,x)(b,j,(x+u(y-x)-d))}(y) = Q_t(a,b) R_{t+1}(b,j) P(D^a = y - d) \\ P(U^a = u) \quad for\ all\ b, c \in F, i, j \in E, t = 1, \dots, T \tag{2.45}$$

The amount of on-hand inventory x_t and order-up-to level inventory y_t do not have an effect on transition probabilities.

Costs are incurred at each period according to the state (Z_t, Y_t, x_t) and amount of order which is $(y_t - x_t)$. The cost at time t is denoted by $c_{(a,i,x,y)}$ where a is the real process, i is the observed process, x is the amount of inventory in hand and y is order up to level. Then, $c_{(a,i,x,y)}$ is computed by the equation (10). (Notations in the dynamic programming model are used for computing costs):

A model can be adapted to find minimum expected discounted cost. It is expressed as:

$$minimize \sum_{a\in F}\sum_{i\in E}\sum_{x\in X} p_0 V_{(a,i,x,0)}$$

$$subject\ to$$

$$V_{(a,i,x,t)} = \sum_{y=x}^{x+K} \alpha_{(a,i,x,y,t)} \Big(c_{(a,i,x,y)} + \sum_{b\in F}\sum_{j\in E}\sum_{d\in D^a}\sum_{u\in U^a} p_{(a,i,x)(b,j,x+u(y-x)-d)}(y)\ V_{(b,j,x+u*(y-x)-d,t+1)}\Big) \tag{2.46}$$

$$for\ all\ (a,b)\in F,(i,j)\in E, x\in X, t=0,1,\dots,T-1$$

$$\sum_{y=x}^{x+K} \alpha_{(a,i,x,y,t)} = 1 \qquad for\ all\ a\in F, x\in X \tag{2.47}$$

$$\alpha_{(a,i,x,y,t)} = \alpha_{(b,i,x,y,t)} \quad for\ all\ a,b\in F, i\in E, x\in X, y\in A, t=1,\dots,T \tag{2.48}$$

$$V_{(a,i,x,T)} = 0 \qquad for\ all\ a\in F, i\in E, x\in X \tag{2.49}$$

$$\alpha_{(a,i,x,y,t)} \geq 0 \qquad for\ all\ a\in F, i\in E, x\in X, y\in\{x,\dots,x+K\}, t=0,\dots.,T \tag{2.50}$$

Note that both α and V are unknown and the above model is a nonlinear program.

The one-period expected total inventory cost at time t plus the expected discounted cost of the next period is computed by the equation (46) by taking the expectation with respect to the demand and yield distributions. Equation (47) ensures that the sum of the probabilities of making all possible orders are equal to 1 for each defined state. The same probability of taking action y for the same observed state i is ensured by equation (49).

CHAPTER 3

COMPUTATIONAL STUDY

In this chapter, computational study is performed to analyze the optimal policies that are obtained from the nonlinear programming model and the dynamic programming model of Arifoğlu and Özekici (2010). To see the effect of the different demand, supply, cost structures and observation levels on optimal policies, three different data sets are used in Arifoğlu and Özekici (2010). So, the comparison of the results of these models is done using these sets and over the results that they report in Arifoğlu and Özekici (2010).

The aim in Arifoğlu and Özekici (2010) is to analyze the optimal decisions for different observation levels and the impact of the demand and supply uncertainties. To perform this analysis, different demand and supply distributions are used in Data Set 1, Data Set 2 and Data Set 3. In the present study, the optimal policies obtained from dynamic programming and nonlinear programming models, are compared in terms of optimal policies and optimal costs. Besides, Data Set 3 is used to analyze the impact of different levels of demand, holding and shortage costs on optimal ordering policies.

3.1 DATA SET 1

In this part, optimal policies are found for different observation levels in which decision maker's precision of observing the real state changes. There are two states of the outside environment Z_t, denoted by $E = \{1,2\}$ where 1 represents *good* and 2 represents *bad* state. The state of the observed process Y_t is $F =$

$\{1,2\}$. The initial probability of the real process is in *good* and *bad* state are $P\{Z_o = 1\} = 0.6$ and $P\{Z_o = 2\} = 0.4$ respectively at time 0. Demand is normally distributed with mean 6 and standard deviation 1 and with mean 4 and standard deviation 1 when the real process is in *good* and *bad* state respectively. Normal distribution is truncated three standard deviations from the mean of the demand and discretized as three point approximation. Random yields are uniformly distributed on the intervals $[0.9,1]$ and $[0.6,0.7]$ for *good* and *bad* states respectively. Transition probability matrix of the real process is

$$Q = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} . \tag{3.1}$$

For the analysis of optimal policies, different observation levels are taken into account through the emission matrix

$$R_i = \begin{bmatrix} 0.45 + 0.05i & 0.55 - 0.05i \\ 0.55 - 0.05i & 0.45 + 0.05i \end{bmatrix}. \tag{3.2}$$

for $i = 0, 1, \dots, 11$. The emission matrix gives the conditional probability of the observed states given that the unobserved states. When $i = 1$, the emission matrix is R_1, and the precision of the observations are at lowest level and observations do not give any information about the state of the real process, every state is equally likely. Precision of the observations increase as observation level (i) increases. R_{11} represents the case where observations are perfect so the conditional probabilities are equal to 1.

The parameters are

- Purchasing cost (c) = 0.5
- Holding cost (h) = 2
- Shortage cost (p) = 10
- Discount factor (α) = 0.9
- Number of periods (T) = 2.

- Capacity of Order (K) = 5

Because main focus in this part is to observe the effect of the precision of observations on optimal policies, cost parameters are assumed to be same for each state of the observed process. If there is random yield in supply, the optimal order-up-to level is characterized by two numbers $(s_0^{\pi_0,i}, S_0^{\pi_0,i})$ when there are two periods remaining at time 0. The vector π_0 is the conditional distribution of the real process Z_0 given $\boldsymbol{Y_0} = (Y_0, Y_1)$. If the initial inventory level is below $s_0^{\pi_0,i}$, it is optimal to order the capacity $K = 5$. If it is above $S_0^{\pi_0,i}$, then order should not be made. When the initial inventory level is between these points, optimal order-up-to level $y_0^{\pi_0,i}$ is greater than $S_0^{\pi_0,i}$ and it is nonincreasing with the inventory level. The structure of the optimal policy is the same when there is one period remaining at time 1. This type of optimal policy is defined by state-dependent modified inflated base-stock policy. It is proved that this policy is optimal for single and multiple periods.

Table 3.1.1 Optimal threshold levels for different emission matrices (Results of Dynamic Programming Model)

Emission	Time 0				Time 1			
	$s_0^{\pi_0,1}$	$S_0^{\pi_0,1}$	$s_0^{\pi_0,2}$	$S_0^{\pi_0,2}$	$s_0^{\pi_1,1}$	$S_0^{\pi_1,1}$	$s_0^{\pi_1,2}$	$S_0^{\pi_1,2}$
R_1	2.9	7.3	2.9	7.3	1.9	6.4	1.9	6.4
R_2	2.9	7.3	2.9	7.3	1.9	6.5	1.8	6.3
R_3	2.9	7.3	2.9	7.3	1.9	6.5	1.8	6.2
R_4	2.9	7.3	2.9	7.3	1.9	6.6	1.8	6.1
R_5	2.9	7.3	2.9	7.3	2	6.6	1.8	6
R_6	2.9	7.3	2.9	7.3	2	6.7	1.7	5.8
R_7	2.9	7.3	2.9	7.3	2	6.7	1.7	5.6
R_8	2.9	7.3	2.9	7.3	2	6.7	1.7	5.4
R_9	2.9	7.3	2.9	7.3	2	6.8	1.6	5.2
R_{10}	2.9	7.3	2.9	7.3	2	6.8	1.6	5
R_{11}	2.9	7.3	2.9	7.3	2	6.8	1.5	4.8

Table 3.1.1 shows optimal threshold values that are obtained from dynamic programming model. Inventory levels is incremented with 0.1 between -5 and 10. and the order-up-to levels are between -5 and 15 with the same level of increments. The length of the intervals are the same with the Arifoğlu and Özekici (2010) to compare the results appropriately. MATLAB is used for solving dynamic program. The threshold levels are same except $S_0^{\pi_0,i}$ values that are 7.5 in their study.

In our nonlinear programming model, initial inventory level is discretized by 76 equidistant points between -5 and 10 for the approximation of normal distribution. More points for the inventory level cannot be considered due to the capacity limitation of GAMS. We assume that the yield is not random ($U = 1$) in our computations. The problem is also solved via dynamic programming with the same precision of the inventory levels. Table 3.1.2 shows optimal threshold levels of both models at time 0.

Table 3.1.2 Optimal threshold levels for different emission matrices at time 0 using 0.2 increments

Emission	Nonlinear Model				Dynamic Programming Model			
	s_0^1	S_0^1	s_0^2	S_0^2	$s_0^{\pi_0,1}$	$S_0^{\pi_0,1}$	$s_0^{\pi_0,2}$	$S_0^{\pi_0,2}$
R_1	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_2	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_3	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_4	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_5	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_6	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_7	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_8	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_9	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_{10}	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8
R_{11}	1.8	6.8	1.8	6.8	1.8	6.8	1.8	6.8

Figure 3.1.1 shows optimal order-up-to levels, which are obtained from our nonlinear program, for the inventory levels between -5 and 10 for the emission matrixes $R_1, R_4, R_7, R_{11.}$ When there are two periods remaining at time 0, optimal order-up-to level is characterized by two numbers ($s_0^i = 1.8,\ S_0^i = 6.8$) that are independent of the observation i in this particular case . The structure of the optimal policy is the same as the solution of the dynamic programming program. The optimal threshold levels are presented in Table 3.1.1 for different observation levels. The optimal policy from dynamic program is characterized by two numbers ($s_0^{\pi_0,i}, S_0^{\pi_0,i}$) when there are two periods remaining at time 0. If $\in \left[s_0^{\pi_0,i}, S_0^{\pi_0,i}\right]$, it is optimal to order to the amount $S_0^{\pi_0,i}$, if x is less than $s_0^{\pi_0,i}$ it is optimal to order $x + K$. Otherwise, no order is made.

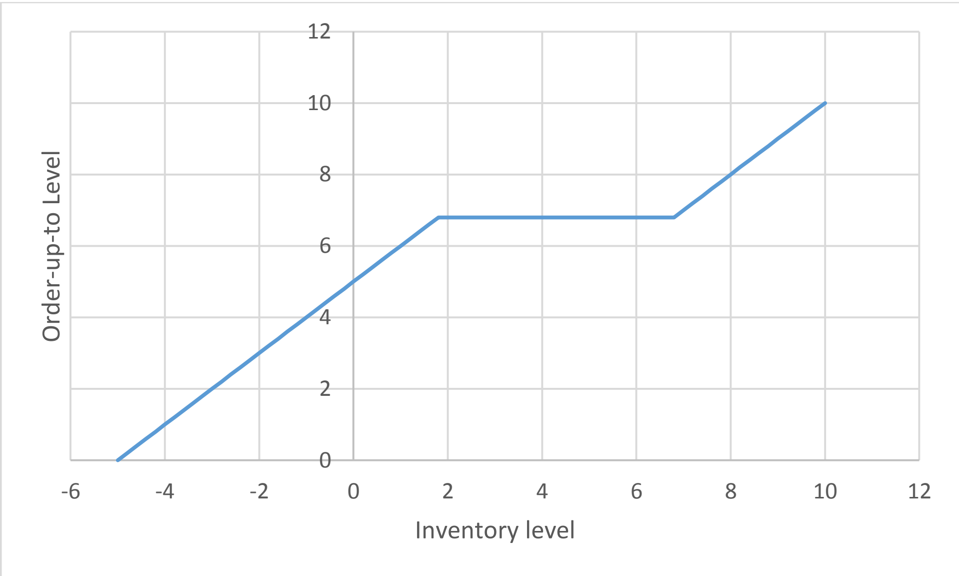

Figure 3.1.1 Optimal order-up-to level at time 0 for R_1, R_4, R_7, R_{11} and Observed State = 1, 2

At time 0, the initial distribution of the unobserved process (Z_0) is known so observing *good* or *bad* states does not affect the transition probabilities. Cost parameters are the same for each state of the observed process. Therefore, in the results of our nonlinear model, the critical levels (s_0^i, S_0^i) are the same for both of the observed states. They are also the same for each emission matrix in the solutions that need not hold in general. The results of dynamic programming and nonlinear program are the same.

When the outside environment Y_1, is completely unobservable with the emission matrix R_1, optimal policies are the same whether *good* or *bad* state is observed. This results from the fact that the observations do not give any information about the state of the real process Z_t. It can be observed from the Table 3.1.3 that threshold values are nondecreasing with the observation level when the state is *good*. When the observation is *good* ($i = 1$) and the emission matrix is R_1 where the environment is completely unobservable, the threshold levels are ($s_1^1 = 1.6, S_1^1 = 6.6$).When the environment is completely observable and the emission matrix is R_{11}, the threshold values are ($s_1^1 = 1.8, S_1^1 = 6.8$). In the former, when the observation is *good*, the probability of real state being *good* is only 0.5 where in the latter that probability is 1. Since *good* state means higher mean demand, the thresholds increase. When $i = 2$ is observed, the observation is *bad* and the threshold levels are ($s_1^2 = 1.6, S_1^2 = 6.6$) for the emission matrix R_1 where the environment is completely unobservable. When the emission matrix is R_{11}, the environment is completely observable and the threshold values are ($s_1^2 = 0, S_1^2 = 4.8$). In the latter case, real process Z_t is perfectly observable which means if the decision maker observes the state as *good (bad)* then it is $good$ (bad) with probability 1.

Table 3.1.3 Optimal threshold levels for different emission matrices at time 1

Emission	Nonlinear Model				Dynamic Model			
	s_1^1	S_1^1	s_1^2	S_1^2	$s_1^{\pi_0,1}$	$S_1^{\pi_0,1}$	$s_1^{\pi_0,2}$	$S_1^{\pi_0,2}$
R_1	1.6	6.6	1.6	6.6	1.6	6.6	1.6	6.6
R_2	1.6	6.6	1.6	6.6	1.6	6.6	1.6	6.6
R_3	1.8	6.8	1.4	6.4	1.8	6.8	1.4	6.4
R_4	1.8	6.8	0.4	5.4	1.8	6.8	0.4	5.4

Table 3.1.3 (Continued)

R_5	1.8	6.8	0.4	5.4	1.8	6.8	0.4	5.4
R_6	1.8	6.8	0.2	5.2	1.8	6.8	0.2	5.2
R_7	1.8	6.8	0.2	5.2	1.8	6.8	0.2	5.2
R_8	1.8	6.8	0	5	1.8	6.8	0	5
R_9	1.8	6.8	0	5	1.8	6.8	0	5
R_{10}	1.8	6.8	0	5	1.8	6.8	0	5
R_{11}	1.8	6.8	0	4.8	1.8	6.8	0	4.8

Figure 3.1.2, Figure 3.1.3, Figure 3.1.4 and Figure 3.1.5 show optimal order-up-to levels for the inventory levels between 0 and 10 for the emission matrices R_1, R_4, R_7, R_{11} respectively at time 1 when there is one period remaining. The optimal order-up-to levels are represented by $y*(1,1)$ and $y*(2,1)$ in the graphs when the observed states is *good* (1) and *bad* (2) respectively. The results of dynamic programming and nonlinear program are exactly the same. These figures demonstrate the effect of the accuracy of the observations about the real state on the optimal order-up-to levels as explained above.

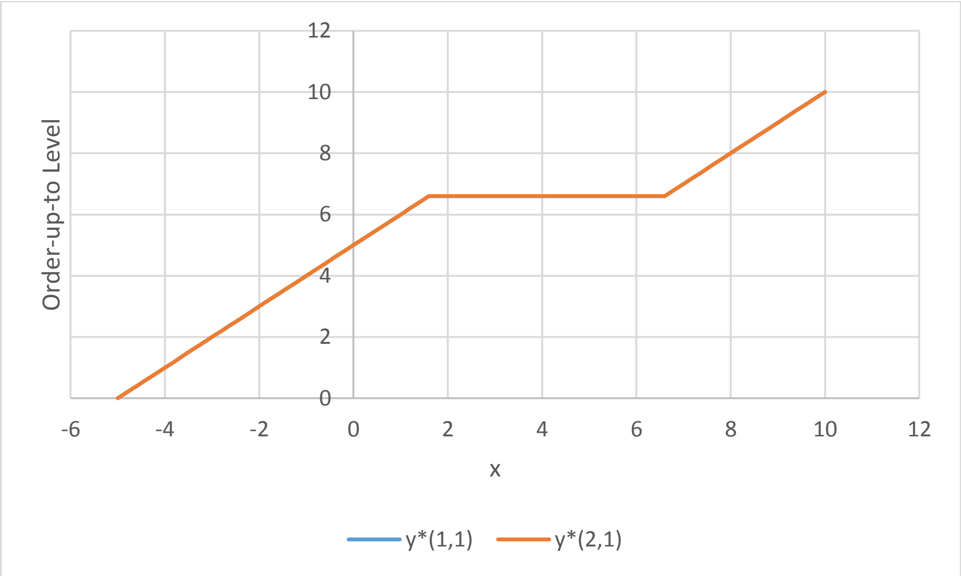

Figure 3.1.2 Optimal order-up-to levels at time 1 for R_1 and Observed State = 1, 2

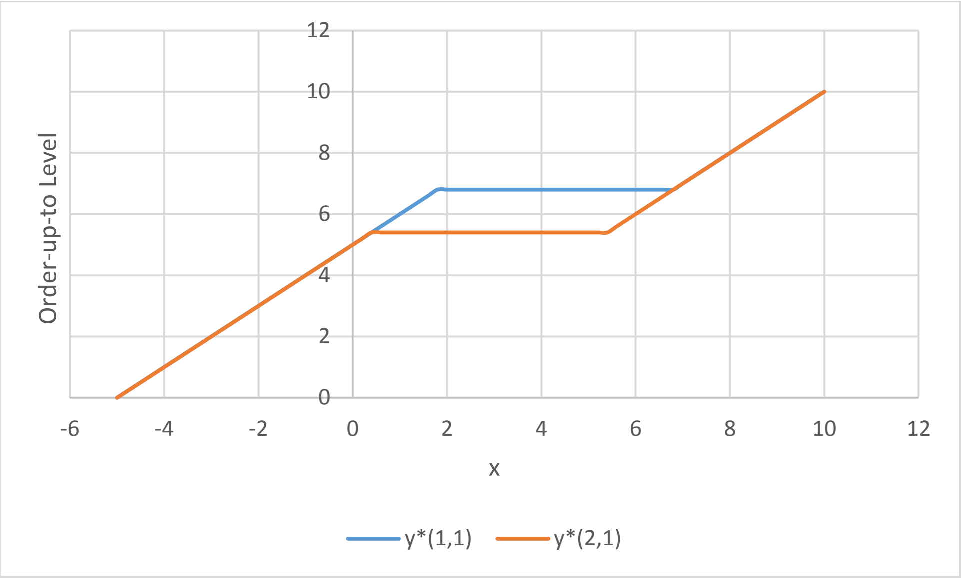

Figure 3.1.3 Optimal order-up-to level at time 1 for R_4 and Observed State = 1, 2

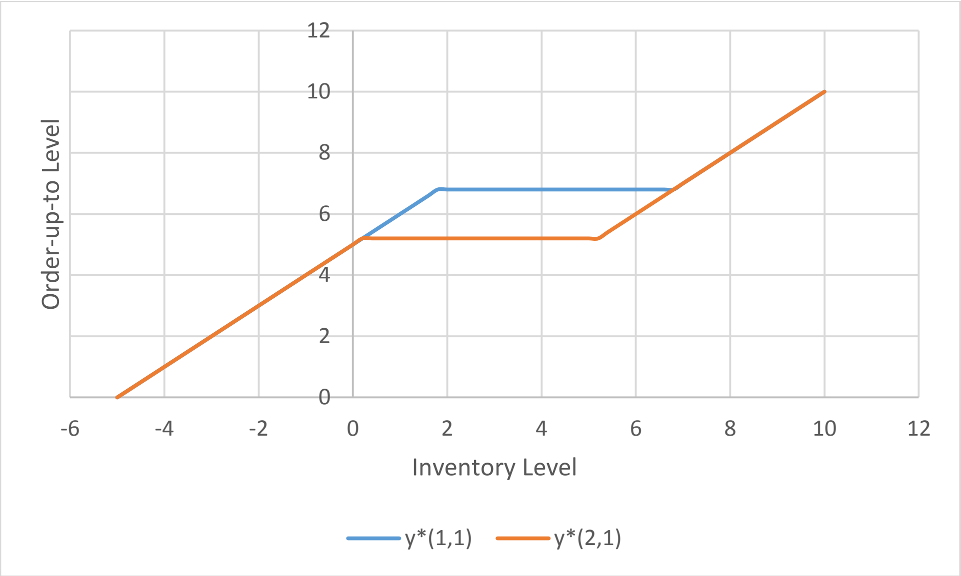

Figure 3.1.4 Optimal order-up-to level at time 1 for R_7 and Observed State = 1, 2

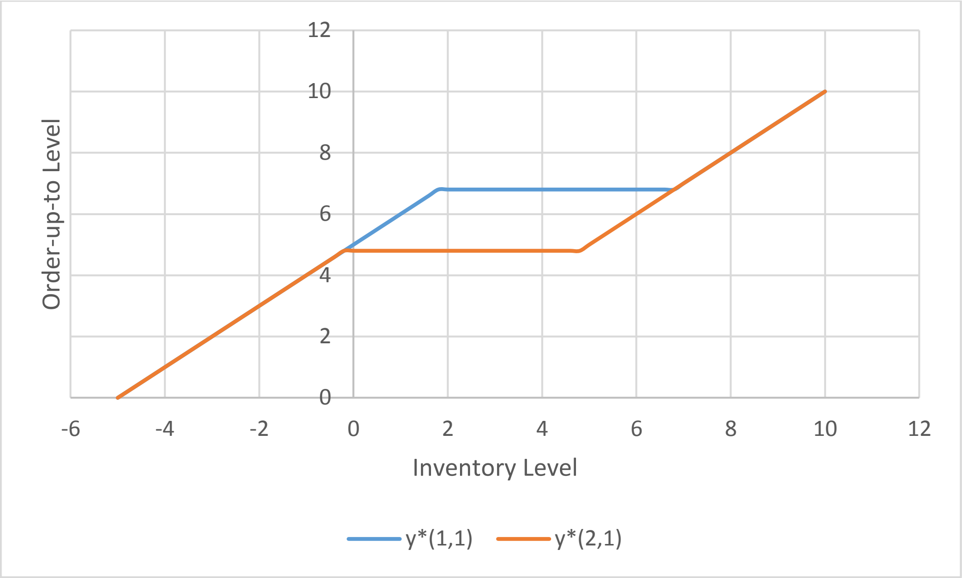

Figure 3.1.5 Optimal order-up-to level at time 1 for R_{11} and Observed State = 1, 2

As Table 3.1.4 shows that optimal cost functions at time 0 are decreasing in the precision of the observations for $i = 1,2$. This results from the fact that the probability of observing different state than the real state is decreasing. Therefore, the cost due to the difference between real and observed states decreases. Figure 3.1.6 shows the optimal expected cost against the initial inventory. The convex behavior is due to the excess or shortage created by the initial inventory. Obviously, starting with the inventory level close to the optimal provides the minimum optimal cost for every emission level. The value of the accurate observation is higher around the “best” initial inventory level. If the initial inventory is too low $(x = 0)$ or too high $(x = 10)$, the saving due to accurate observation is low.

Table 3.1.4 Expected total optimal costs values for different initial inventory levels for R_1, R_4, R_7, R_{11}

Initial Inventory Level	Emission Matrices			
	R_1	R_4	R_7	R_{11}
0	21.497	21.486	21.360	20.980
2	12.830	12.777	12.417	11.570
4	11.830	11.783	11.439	10.614
6	10.830	10.783	10.439	9.614
8	11.863	11.077	10.583	9.543
10	14.190	14.102	13.666	12.852

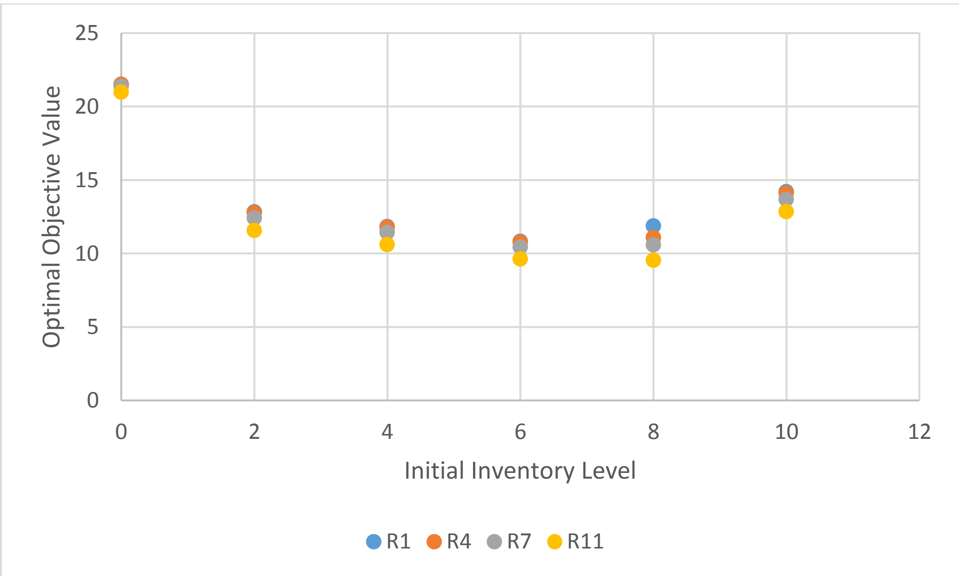

Figure 3.1.6 Optimal objective values for different initial inventory levels for R_1, R_4, R_7, R_{11}

3.2 DATA SET 2

In this part, the distribution of the demand is Poisson and the yield distribution is uniform. In addition, Poisson distribution is truncated three standard deviations from its mean for the demand distributions. Parameters of both of these distributions depend on the state of the observed process Y_t. Costs are different for each state of the observed process. All other parameters are the same as in Data Set 1. Integer values are considered for the inventory levels because demand is distributed as Poisson and inventory level is rounded to the nearest integer after some proportion of the order received. Uncertainty in demand and supply is measured by the coefficient of variation.

The cost parameters are

- Purchasing costs: $c_1 = 0.5,\ c_2 = 1$
- Holding costs: $h_1 = 2,\ h_2 = 4$
- Shortage costs: $p_1 = 10,\ p_2 = 20.$

Parameters of the Poisson and Uniform distribution are

	Good	Bad
Mean Demand (λ)	4	2
Supply Interval $[a, b]$	[0.5,0.8]	[0.4,0.5].

The results, which are obtained from dynamic programming program and nonlinear program are stated in Table 3.2.1, are considered for the initial inventory level between 0 and 10 at time 0. MATLAB is used for solving dynamic program.

In the nonlinear program, initial inventory level is considered as integer points between 0 and 10. GAMS is used for solving the nonlinear program. We assume that there is no proportional yield ($U = 1$) in our computations.

Figure 3.2.1 shows the optimal order-up-to levels at time 0. When there are two periods remaining at time 0, optimal order-up-to level is characterized by two numbers ($s_0^i = 0,\ S_0^i = 5$) for $i = 1,2$. The structure of the optimal policy is the same as the solution of the dynamic programming program. The optimal policy is characterized by two numbers ($s_0^{\pi_0,i}, S_0^{\pi_0,i}$) when there are two periods remaining at time 0. If x ∈ $\left[s_0^{\pi_0,i}, S_0^{\pi_0,i}\right]$, it is optimal to order to the amount $S_0^{\pi_0,i}$, if x is less than $s_0^{\pi_0,i}$ it is optimal to order $x + K$. Otherwise, no order is made. This solution ($s_0^i = 0,\ S_0^i = 5$), is the same for all observation levels.

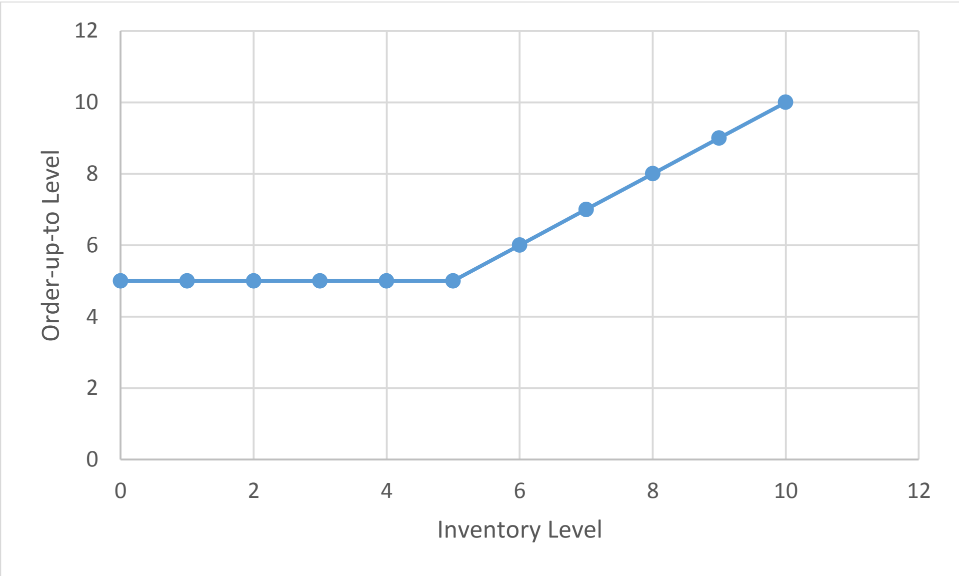

Figure 3.2.1 Optimal order-up-to levels at time 0 for R_1, R_4, R_7, R_{11} and Observed State = 1, 2

When the outside environment Y_1, is completely unobservable with the emission matrix R_1, optimal policies are the same whether *good* or *bad* state is observed. This results from the fact that the observations do not give any information about the state of the real process Z_t. It can be observed from the Table 3.2.1 that threshold values are nondecreasing with the observation level when the state is *good*. When the observation is *good* ($i = 1$) and the emission matrix is R_1 where the environment is completely unobservable, the threshold levels are ($s_1^1 =$

$0, S_1^1 = 5)$. When the environment is completely observable and the emission matrix is R_{11}, the threshold values are $(s_1^1 = 1, S_1^1 = 6)$. In the former, when the observation is *good*, the probability of real state being *good* is only 0.5 where in the latter that probability is 1. Since *good* state means higher mean demand, the thresholds increase. When $i = 2$ is observed, the observation is *bad* and the threshold levels are $(s_1^2 = 0, S_1^2 = 5)$ for the emission matrix R_1 where the environment is completely unobservable. When the emission matrix is R_{11}, the environment is completely observable and the threshold values are $(s_1^2 = 0, S_1^2 = 3)$. In the latter case, real process Z_t is perfectly observable which means if the decision maker observes the state as *good (bad)* then it is $good$ (bad) with probability 1.

Table 3.2.1 Optimal threshold levels for different observation levels at time 1

Emission	Nonlinear Model				Dynamic Programming Model			
	s_1^1	S_1^1	s_1^2	S_1^2	$s_1^{\pi_1,1}$	$S_1^{\pi_1,1}$	$s_1^{\pi_1,2}$	$S_1^{\pi_1,2}$
R_1	0	5	0	5	0	5	0	5
R_2	0	5	0	4	0	5	0	4
R_3	0	5	0	4	0	5	0	4
R_4	0	5	0	4	0	5	0	4
R_5	0	5	0	4	0	5	0	4
R_6	0	5	0	4	0	5	0	4
R_7	0	5	0	4	0	5	0	4
R_8	0	5	0	3	0	5	0	3
R_9	0	5	0	3	0	5	0	3
R_{10}	0	5	0	3	0	5	0	3
R_{11}	1	6	0	3	1	6	0	3

Figure 3.2.2, Figure 3.2.3, Figure 3.2.4 and Figure 3.2.5 show optimal order-up-to levels for the inventory levels between 0 and 10 for the emission matrices R_1, R_4, R_7, R_{11} respectively at time 1 when there is one period remaining. The optimal order-up-to levels are represented by $y*(1,1)$ and $y*(2,1)$ when the observed states is $good$ (1) and bad (2) respectively. The results of dynamic programming and nonlinear program are exactly the same. These figures

demonstrate the effect of the accuracy of the observations about the real state on the optimal order-up-to levels as explained above

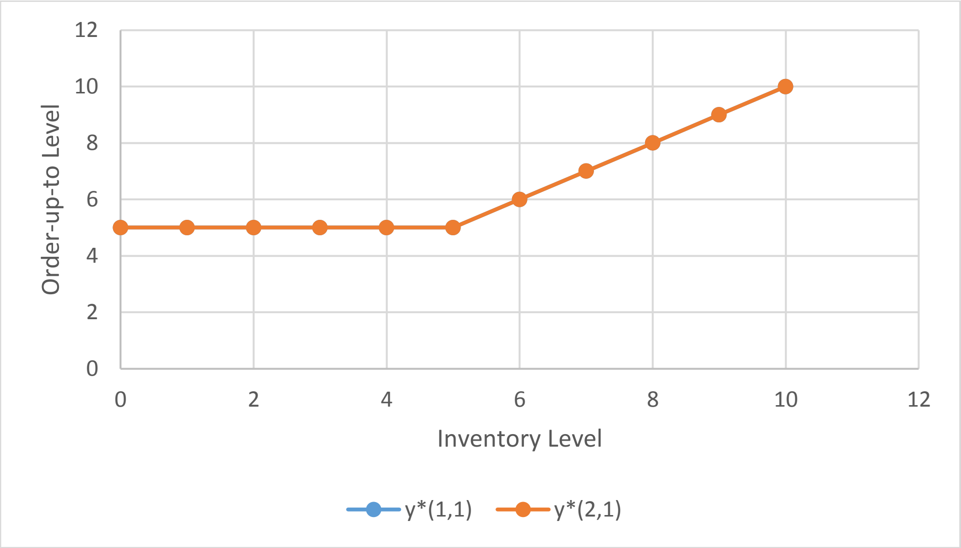

Figure 3.2.2 Optimal order-up-to levels at time 1 for R_1 and Observed State = 1, 2

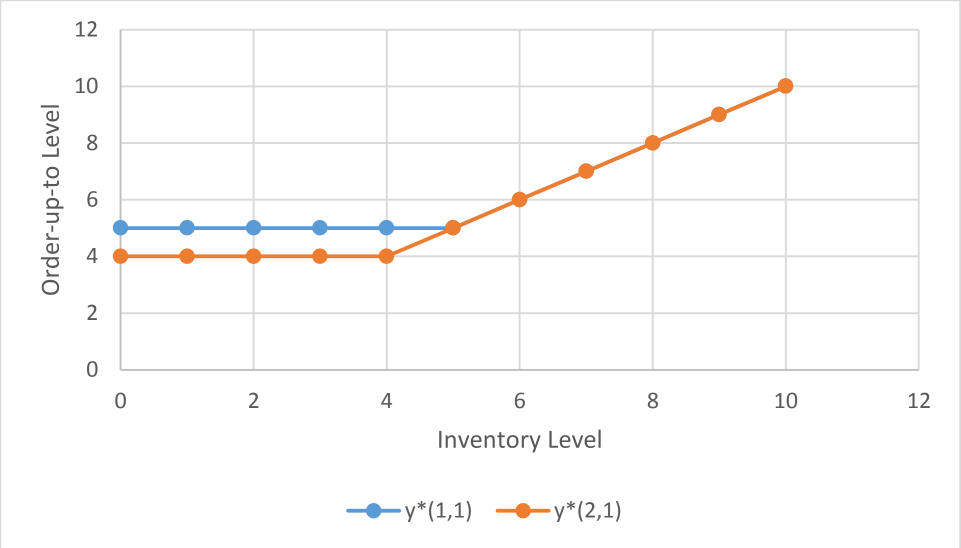

Figure 3.2.3 Optimal order-up-to levels at time 1 for R_4 and Observed State = 1, 2

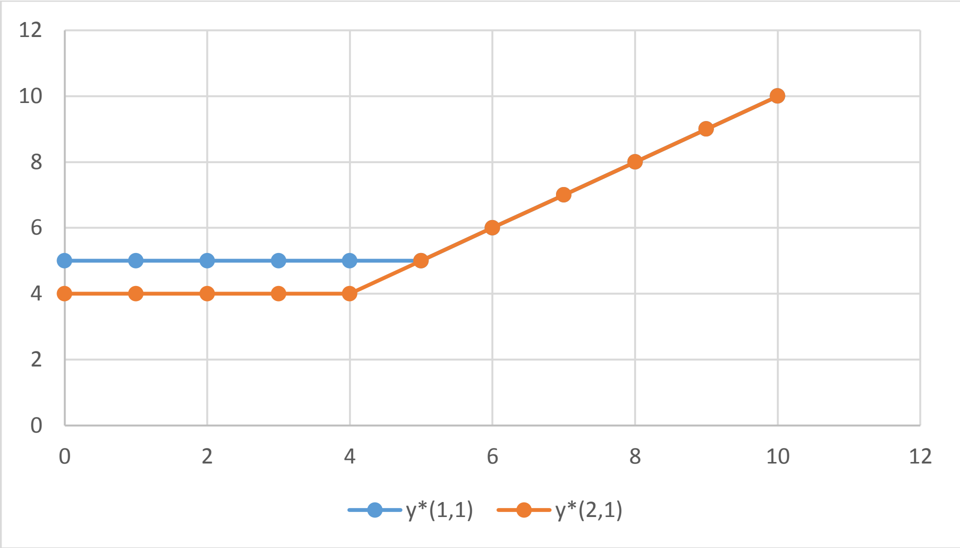

Figure 3.2.4 Optimal order-up-to levels at time 1 for R_7 and Observed State = 1, 2

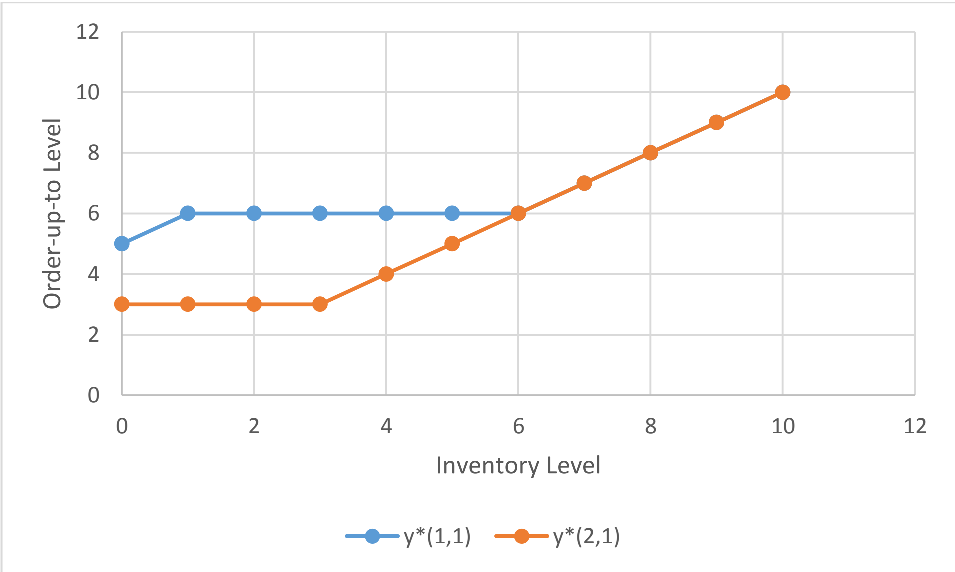

Figure 3.2.5 Optimal order-up-to levels at time 1 for R_{11} and Observed State = 1, 2

3.3 DATA SET 3

In this part, the effect of different demand levels and cost parameters on optimal policies are analyzed. Different levels of demand, holding cost and shortage cost are considered under three cases. Demand is assumed to be two point discrete. It is assumed that cost parameters are the same for each observed state and the states of the real process, observable process and transition probabilities are the same as in Data Set 1. The problem is solved by both dynamic programming model and nonlinear programming model and the same results are obtained. The multiple-period optimal ordering policy is a state-dependent modified base-stock policy for all of the cases that are considered. This policy structure is characterized by two numbers $(s_0^i,\ S_0^i)$ when there are two periods remaining at time 0. If $x \in [s_0^i,\ S_0^i]$, it is optimal to order to the amount S_0^i, if x is less than s_0^i it is optimal to order capacity. Otherwise, no order is made.

The parameters of the problem are

- Purchasing cost (c) = 0.5
- Discount factor (α) = 0.9
- Number of periods (T) = 2
- Capacity of Order (K) = 5.

Figure 3.3.1 shows the optimal order-up-to levels for different mean demands. The mean of the demand increases from Case 1 to Case 3 so there is more risk of shortage if the same ordering decision is made for different cases. The optimal threshold levels are $(s_0^1 = 0,\ S_0^2 = 5)$, $(s_0^1 = 4,\ S_0^2 = 9)$, $(s_0^1 = 10,\ S_0^2 = 15)$ for Case 1, 2, 3 respectively. It is reasonable that the optimal order-up-to levels from Case 1 to Case 3 increase as the mean of the demand increases to minimize total expected cost.

The effect of different holding costs on optimal policies are considered in Figure 3.3.2. The holding cost increases from Case 1 to Case 3 so more holding cost is incurred for the same level ordering decision. Therefore, the optimal order-up-to levels decrease from Case 1 to Case 3 as expected. The optimal threshold levels are $(s_0^1 = 6,\ S_0^2 = 11)$, $(s_0^1 = 4,\ S_0^2 = 9)$, $(s_0^1 = 2,\ S_0^2 = 7)$ for Case 1, 2, 3 respectively.

Figure 3.3.3 illustrates the optimal order-up-to levels for different shortage costs. Shortage costs increase from Case 1 to Case 3 so more shortage cost is incurred for the same ordering level in different cases. Therefore, the optimal order-up-to levels increase from Case 1 to Case 3 as expected. The optimal threshold levels are $(s_0^1 = 4,\ S_0^2 = 9)$, $(s_0^1 = 6,\ S_0^2 = 11)$, $(s_0^1 = 8,\ S_0^2 = 13)$ for Case 1, 2, 3 respectively.

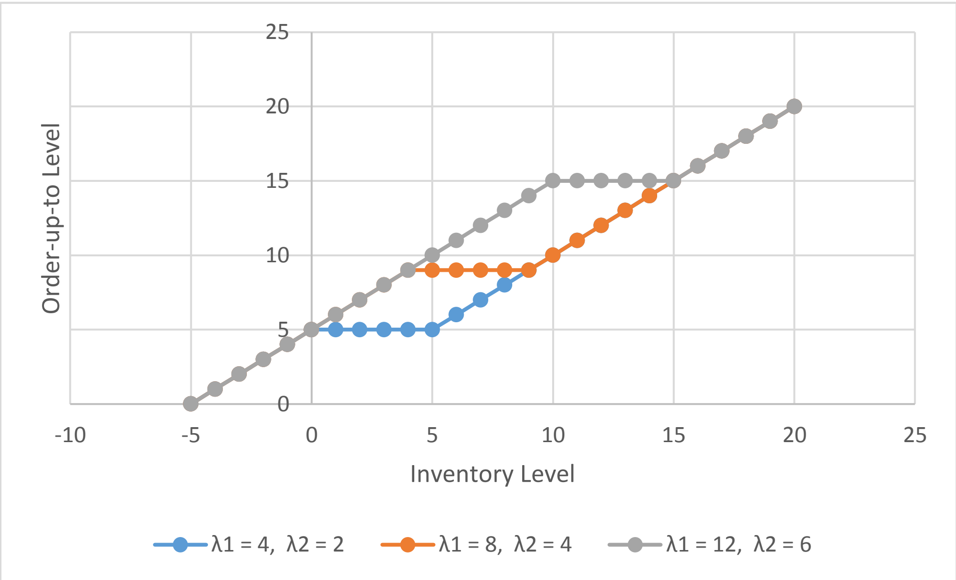

Figure 3.3.1 Optimal order-up-to level at time 0 and Observed State = 1, 2 for different demand levels ($h = 2, p = 10$)

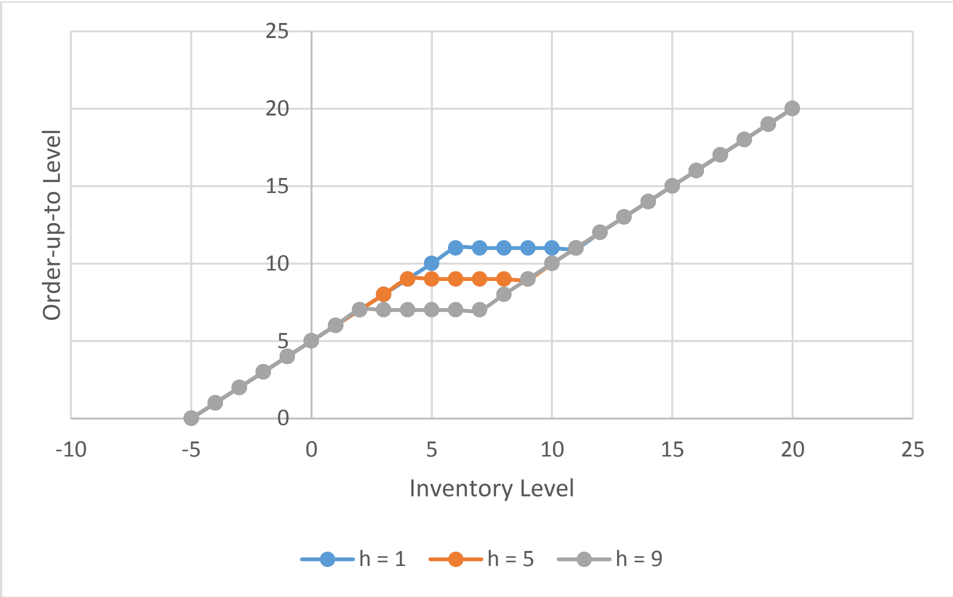

Figure 3.3.2 Optimal order-up-to level at time 0 and Observed State = 1, 2 for different holding costs ($\lambda_1 = 8,\ \lambda_2 = 4$)

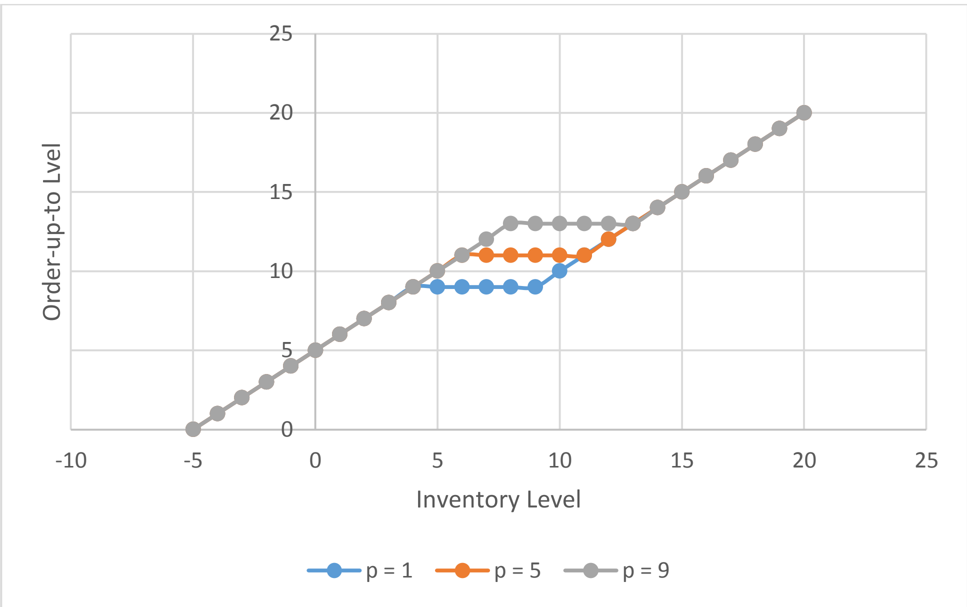

Figure 3.3.3 Optimal order-up-to level at time 0 and Observed State = 1, 2 for different shortage costs ($\lambda_1 = 8,\ \lambda_2 = 4$)

3.4 CONCLUSION

In this study, a finite horizon inventory problem with random demands that are generated by a partially observable Markov environment posed in Arifoğlu and Özekici (2010) is taken. There is an observable process that is probabilistically related to the real process. Arifoğlu and Özekici (2010) constructed a dynamic program to minimize the total cost over the finite horizon using POMDP methodology. They showed that the optimal ordering policy depends on the probability distribution over the states and it is control limit type. They provide numerical examples for two periods.

Serin and Avşar (1997) constructed a nonlinear problem that provides approximate simpler solutions to the POMDP problems. The nonlinear programming model for the above inventory problem is constructed and the results are compared for several cases. The optimal policies that are found by the dynamic programming model and nonlinear programming model are exactly the same for all of the data sets so there is no difference in the objective function values. The structure of the optimal policy is state-dependent modified base-stock policy as expected.

The effect of the precision of the observations on the optimal policies are considered by the different emission matrices with different observation levels from the completely unobservable case to the completely observable case. The optimal objective function values are decreasing as the precision of the observations increase as expected in all of the data sets.

The effect of the initial inventory on the optimal expected cost for different observation levels are analyzed. The value of the accurate observation is higher around the “best” initial inventory level. If the initial inventory is too low or too high, the saving due to accurate observation is low.

The effects of some model parameters such as the mean demand, the holding and shortage costs on the optimal order up-to-levels are also analyzed.

REFERENCES

Arifoğlu, K., Özekici, S., 2010. Optimal policies for inventory systems with finite capacity and partially observed Markov-modulated demand and supply processes, European Journal of Operations Research 204, 421-483.

Arifoğlu, K., Özekici, S., 2011. Inventory management with random supply and imperfect information: A hidden Markov model, Int. J. Production Economics 134, 123-137.

F.W Ciarallo, R. Akella, T.E Morton, 1994. A periodic review, production planning model with uncertain capacity and uncertain demand-optimality of extended myopic policies, Management Science, 40, 320-332.

Erdem A., Özekici S., 2002. Inventory models with random yield in a random environment, Int. J. Production Economics, 78, 239-253.

Federgruen A., Zipkin P., 1986.An inventory model with limited production capacity and uncertain demands. Part1: The average-cost criterion, Mathematics of Operations Research, 11, 208-215.

Gallego G., Hu H., 2004. Optimal policies for Production/Inventory systems with finite capacity and Markov-modulated demand and supply processes, Annals of Operations Research, 126, 21-41.

Henig, I.M, Gerchak 1990. Y., The structure of periodic review policies in the presence of random yield. Operations Research 38, 634, 643.

Iglehart D.L, Karlin S., 1962. Optimal policy for dynamic inventory process with nonstationary stochastic demands, Studies in Applied Probability and Management Science, Standord University Poress, Stanford, CA, 127-147.

Karlin S., Fabens A. 1959. The (sS) inventory model under Markovian demand process. Mathematical methods in Social Sciences, Stanford University Press, Stanford, CA.

Özekici S., Parlar M. 1999. Inventory models with unreliable suppliers in a random environment, Annals of Operations Research 91, 123-136

Serin Y., Avşar Z.M, 1997. Markov Decision Processes with Restricted Observations: Finite Horizon Case, Naval Research Logistics, 44, 439-456.

Smallwood R.D, Sondik E.J, 1973. Optimal control of partially observable processes over finite horizon. Operations Research 21, 1071-1088.

Song J.S, Zipkin P., 1993. Inventory control in a fluctuating demand environment, Operations Research, 41, 351-370

Wang Y., Gerchak Y., Periodic review production models with variable capacity, random yield and uncertain demand, Management Science, 42, 130-137