# STOCHASTIC INVENTORY CONTROL WITH PARTIAL DEMAND OBSERVABILITY

A Thesis
Presented to
The Academic Faculty

by

Olga L Ortiz

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
April 2008

# STOCHASTIC INVENTORY CONTROL WITH PARTIAL DEMAND OBSERVABILITY

Approved by:


Alan L. Erera, Advisor
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Chelsea C. White, III, Advisor
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Julie L. Swann
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Paul Griffin
School of Industrial and Systems Engineering
*Georgia Institute of Technology*

Soumen Ghosh
College of Management
*Georgia Institute of Technology*



Date Approved: March 28 2008

*To my husband,*

*Juan C. Morales,*

*And our son*

*Daniel*

# ACKNOWLEDGEMENTS

First of all I would like to extend my deepest thanks to my advisors Dr. Alan L. Erera and Dr. Chelsea C. White III for their guidance, their support and mostly their understanding. Without them this dissertation would not exist. I want to thank my committee members Dr. Julie Swann, Dr. Paul Griffin and Dr. Soumen Ghosh for their valuable comments.

I want to thank my family for their support. I also want to thank my friends at Georgia Tech who made my life in Atlanta enjoyable. Finally I want to extend my most sincere thank to my husband Juan for his support, understanding and for always believing in me.

# TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# SUMMARY

This dissertation focuses on issues associated with the value of information in models of sequential decision making under uncertainty. All of these issues are motivated by inventory management problems. First, we study the effect of the accuracy of inventory counts on system performance when using a zero-memory controller in an inventory system that is modeled as a partially observed Markov decision process (POMDP). We derive conditions for which improving the accuracy of inventory counts will either (i) improve system performance, (ii) degrade system performance or (iii) will not affect system performance. With a computational study, we determine the range of profitability impacts that result from inaccurate inventory counts when using reasonable zero-memory control policies.

Second, we assess the value of demand observation quality in an inventory system with Markovian demand and lost sales. Again, the POMDP serves as a problem model, and we develop computationally tractable suboptimal algorithms to enable the computation of effective lower bounds on system profitability when demand observations are noise-corrupted. We then extend our results to consider the effects that product substitution has on system performance. We show that systems with low demand variability, high holding cost levels, and high levels of substitution benefit more from demand observability than systems with high demand variability, low holding cost levels, and low levels of substitution.

Third, to enhance our understanding of sequential inventory control with substitutable products, we analyze a two-item inventory problem with known deterministic primary demand, but stochastic one-way substitution. We model this problem as a MDP and show that a decision rule that minimizes the single period cost function, when applied at every decision epoch over the infinite horizon, is an optimal policy for the infinite horizon problem. A definition of increased substitutability is presented, and it is shown that increased substitutability never increases optimal expected total discounted cost.

# CHAPTER I

# INTRODUCTION

Sequential decision making under uncertainty is a complex and fascinating area of study that has captured the interest of many researchers. Techniques based on Markov Decision Processes (MDP) and Stochastic Programming, among others, have been proposed as alternative ways to address the problem of making sequential decisions in a stochastic environment. A very common assumption for most of these approaches is that random realizations of the associated stochastic process can be observed with perfect accuracy. However, in many realistic problem settings, such as the ones described and studied in this dissertation, perfect state observation quality may not exist. Decision makers need to therefore identify strategies that deal with both the stochastic environment as well as with inaccurate observations of the actual underlying process.

This dissertation focuses primarily on problems of sequential decision making under uncertainty where the realization of the underlying stochastic process is partially observed. The models we use are the MDP and the Partially Observed Markov Decision Process (POMDP), an extension of the traditional MDP that explicitly models inaccurate state observations. We propose general models and algorithms and derive results that have a potentially wide range of applications. Inventory systems represent our primary motivation, where we consider cases in which realized demand and inventory levels are partially observed.

Much of our research is motivated by the common practice of using traditional models that assume perfect observation of the underlying stochastic process in contexts where only noise-corrupted observations are available to the decision maker. These noise-corrupted observations are then used in the process of estimating model inputs or applying policies obtained by solving the corresponding model with perfect state observability. For instance, it is common in retailing applications to use sales as true observations of demand, in order

to estimate demand parameters of inventory models that are then solved to define order quantities, reorder points, etc. In this dissertation we examine the validity of this approach and aim at assessing the value of improving the accuracy of the corrupted observations. Note that, most efforts undertaken by the research community focus on traditional approaches that assume perfect observability; comparatively, very limited literature exists on the topic of decision making for partially observed stochastic processes. In this research we make several contributions to the latter area of study.

It seems reasonable that decision makers might be willing to pay an extra cost in order to improve the accuracy of their observations if this indeed leads to more profitable decisions when applying a traditional decision model. In Chapter 2 we determine if decisions made with traditional models, assuming that corrupted observations are true observations, improve as the accuracy of the observations also improve. We show that this intuitive result is not always true; *i.e.*, in general improving the quality of the observation does not necessarily lead to better solutions. We also identify conditions under which improved observability does always lead to better decisions. More specifically, Chapter 2:

- Investigates the impact of improving state observation quality on policies that are a function of only the most recent observation of the POMDP (zero-memory policy),
  - Providing conditions for a zero-memory policy to be adaptive, meaning that investing in improving the quality of the state observation will not degrade the performance of the policy;
  - Determining the existence of non adaptive zero-memory policies; and
  - Providing conditions in which improving state observation quality will not affect system performance when using a given zero-memory policy;
- Develops a POMDP decision model for a single item inventory problem with $i.i.d$ demand and inaccurate inventory counts; and
- Provides a computational study assessing the value of improving inventory accuracy

when using a zero-memory policy that is equivalent to an optimal policy of the Completely Observed Counterpart (COC) of the POMDP, where the COC of the POMDP is the MDP that results when observations are assumed to be perfect.

We show in Chapter 2 that the expected total discounted reward accrued over the infinite horizon for a zero-memory policy, $v$, is the inner product of the current state probability mass vector $x$ and a vector $\gamma$ , *i.e.*, $v = x\gamma$ , where $\gamma$ only depends on the most recent state observation. Further, $\gamma$ can be represented as a power series in $\epsilon$, where $\epsilon$ is the probability of a state observation error. Thus, for small $\epsilon$, $v$ can be approximated by $x(\alpha^0 + \epsilon\alpha^1)$, which implies that whether or not improved state observation quality improves system performance is dependent on the signs of the scalar elements of the vector $\alpha^1$. We show that if $\gamma$ is generated by an optimal policy for the COC, then all scalar elements of $\alpha^1$ are non-positive; therefore, decreasing $\epsilon$, *i.e.*, improving state observation quality, does not decrease $v$, *i.e.*, does not degrade system performance. We show by example that a near-optimal COC zero-memory policy can produce a vector $\alpha^1$ with all positive scalar elements. Thus, it is possible that what appears to be a good suboptimal design can degrade system performance, given improved state observation quality.

In Chapter 2 we also investigate, by means of a computational study of an inventory system where inventory level observations are noise corrupted, the value obtained from improved accuracy of inventory counts. We vary three problem characteristics: (1) demand variability measured as the coefficient of variation of the demand process, (2) the contribution of holding cost relative to overall cost and (3) variability of the observation process measured as the maximum absolute error of the observation quantity. Our numerical results indicate that systems with higher inventory holding costs can benefit more from improved inventory accuracy than systems with lower holding cost levels. Demand variability seems to have a significant effect on the potential benefit of improved accuracy of inventory counts. Variability in the observation process has a significant effect on the value of improved inventory accuracy, as one might expect.

Chapter 3 studies the problem of estimating the value of demand observability in a single product periodic review inventory system with lost sales, where the decision maker selects

a replenishment quantity at each decision epoch with the objective of maximizing expected total discounted profit over a finite planning horizon. We assume that the inventory levels at all decision epochs are completely observed but that demand observations may be noise corrupted. Further, demand is modeled using an exogenous Markov chain. In this chapter we develop a methodology to assess the maximum expected increment in system profitability due to improved demand observability. Chapter 3:

- Develops a POMDP decision model for a single product inventory control system with Markovian demand and lost sales, where demand is partially observed via sales and perhaps some other market signal data;
- Presents an algorithm for determining the optimal policy and the optimal expected total discounted profit;
- Develops three computationally attractive heuristic algorithms, the third of which is based on a non-standard sufficient statistic that enables relatively easy software development;
- Develops a methodology for assessing the maximum potential value of improved demand observability; and
- Quantifies the value of improved demand observability via computational study.

In the computational study, we examine the effect of two problem characteristics on the value of demand observability: (1) variability of the demand process and (2) the contribution of holding cost relative to overall cost. We show that the maximum value of observability can vary significantly with these characteristics, from a minimum of 2% to a maximum of nearly 35%. Scenarios with high relative holding costs benefited the most from improved demand observability. In general, we observed that systems with low demand variability benefit more from improved demand observability than systems with higher levels of demand variability.

Chapters 4 and 5 study demand substitution in inventory systems. Chapter 4 extends the methodology for assessing the value of demand observability developed in Chapter 3

to a problem setting with one-way demand substitution between two products. Substitution demand from product 1 to product 2 is only generated during a stockout of product 1. Demand substitution further contributes to hinder true demand observability because sales data for a product may be the result of its true underlying demand process and the substitution demand from the other product. Chapter 4:

- Develops a POMDP decision model for a two product inventory control system with one way substitution and lost sales, where demand is partially observed via sales and perhaps some other market signal data;
- Presents an algorithm for determining the optimal policy and the optimal expected total discounted profit;
- Develops heuristics based on a suboptimal design that provide near-optimal replenishment policies;
- Develops a methodology for assessing the maximum potential value of improved demand observability; and
- Quantifies the value of improved demand observability via computational study.

Numerical results suggest that the potential benefit of improving demand observability is higher in systems with high levels of substitution than in systems with lower demand substitutability levels. The proposed suboptimal design was observed to perform fairly well in systems with similar products and not as well in systems in which product characteristics (most notably, price) are very different.

In Chapter 5 we study the effects of substitution in inventory systems with perfect observability. Specifically we consider the problem of determining an optimal replenishment policy for a two-item inventory system with deterministic primary demand and stochastic one-way substitution demand generated only after a stockout of product 1. Chapter 5:

- Proves the existence of an optimal myopic policy for the infinite horizon problem, a result that motivates an in-depth examination of the single period cost function;

- Develops conditions that guarantee that the minimum of the single period cost function is such that zero replenishment of item 1 is always optimal, and examines these conditions in the context of two substitutability distributions, the uniform and the binomial distributions;
- Develops an algorithm for determining the policy that minimizes the single period cost function;
- Determines the effect of substitutability on the optimal expected cost, and shows that greater substitutability will not increase optimal expected cost; and
- Provides upper and lower bounds on the optimal expected cost.

Finally, we summarize our results and present topics for future research in Chapter 6.

# CHAPTER II

# ADAPTIVITY AND ZERO-MEMORY POLICIES FOR PARTIALLY OBSERVED MARKOV DECISION PROCESSES

## *2.1 Introduction*

For a given zero-memory policy, we investigate the relationship between state observation quality and system performance for a system modeled as a partially observed Markov decision process having reasonably accurate state observations. We present conditions that imply that a zero-memory policy, given improved state observation quality, (i) will improve system performance, (ii) will degrade system performance, or (iii) will not affect system performance. The intent of these results is to provide insights into whether or not investment in improved state observation quality is beneficial.

It is a common assumption that more accurate state observations will result in improved system performance. For example, it is typically assumed that more accurate inventory counts will reduce inventory holding costs and/or the profit loss due to stock outs. We show that this common assumption does not hold in general for zero-memory policies. For example, there are situations when higher quality inventory counts will degrade expected system performance. Thus, the inventory manager, who may want to initially ask "will better performance due to a more accurate inventory count justify the investment needed to improve the quality of the count?", should first ask "are we sure that inventory system performance will improve if the inventory is more accurately counted?".

The intent of this research is to address the question: when state observations are reasonably accurate, under what conditions will a zero-memory policy provide better system performance if given more accurate state observations? Our initial motivation for addressing this question was inventory control. Inventory levels are usually observed accurately, but not perfectly, which justifies our interest in situations where state observation error is small.

Our interest in zero-memory policies is due to the fact that such policies can serve as

good, easily computed sub-optimal designs for the class of models of sequential decision making under uncertainty that serves as the basis for our analysis. We use the infinite horizon, total discounted reward, partially observed Markov decision process (POMDP) as the basis for analysis since the POMDP, an extension of the (standard) MDP, can model noise corrupted, incomplete, or costly observations of the state process. The MDP assumes the system state is perfectly observed without cost and hence is an inadequate model for our investigation.

The superior modeling validity of the POMDP, relative to the MDP, is in contrast to the superior tractability of the MDP, relative to the POMDP. Although not the focus of the research reported in this chapter, the determination of optimal or good sub-optimal policies for the POMDP has been a source of considerable interest. It is well-known that the probability mass vector over the current underlying state, conditioned on all current and former state observations and all former actions, represents a sufficient statistic for the POMDP (e.g., Striebel [49], Astrom [3]). Unfortunately, the state space of this sufficient statistic is uncountable. Smallwood and Sondik (Sondik [47], Smallwood and Sondik [45], and Sondik [48]) were the first to show that the optimal expected reward-to-go function for the finite-horizon POMDP is piecewise linear and concave in this sufficient statistic and hence has a finite representation. Related procedures for determining an optimal policy for the POMDP can be found in White [54], White and Scherer [56], Hansen [17], Lovejoy [30] and Monahan [33]. Cassandra [10] showed that determining an optimal policy for the POMDP is PSPACE-hard, with exact algorithms running in exponential time and polynomial space in the number of state variables and observations, a fact that has motivated a myriad of numerically less taxing, sub-optimal design approaches for the POMDP. These include approaches found in Cassandra [10] and Parr and Russell [38], a genetic algorithmic approach (Lin at el. [28]), grid techniques (e.g., Brafman [9], Bonet [7], Lovejoy [29]), value function approximations (Hauskrecht [18]) reinforcement learning (Jaakkola and at el. [20], Suematsu and Hayashi [50]), factored representations (Sallans [42]), Monte Carlo methods (Thrun [51]), and finite-memory approaches (White and Scherer [57], Meuleau at el. [32], Aberdeen [1]).

In spite of the fact that inventory counts are rarely perfect, the MDP, rather than the POMDP, has served until recently as the basis for analysis for inventory control problems because of, in part, the aforementioned tractability of the MDP, relative to the POMDP. This tractability advantage is further amplified for inventory problems by the optimality of the computationally useful (s, S), continuous review (Q, R), and order-up-to policy structures for large classes of inventory control problems, policy structures that appear to have no counterparts for the POMDP. There has been, however, growing recent interest in inventory control with inaccurate records. Lee and Ozer [27], provide an excellent introduction to this area, from the viewpoint of attempting to understand the benefits that radio-frequency identification (RFID) technology may hold for supply chain inventory management. In the following paragraphs we summarize a sample of research in this area.

Discrepancies between physical inventory levels and inventory records are commonly found in retailing. As reported in Fisher at el. [15], a common practice known as the *zero-balance walk* is used to reconcile these quantities. When inventory records for a stock keeping unit (SKU) reach the zero level, employees walk through the facility to verify a true stockout. Bensoussan at el. [6] formulate and analyze a zero-balance walk POMDP model by assuming that an accurate inventory count is only available when its level is zero. Nonzero inventory counts are assumed to be known only in distribution. The paper develops and analyzes a periodic review inventory control model in this setting, where unsatisfied demand each period is lost and the objective is to minimize total discounted cost over an infinite horizon. An approximately optimal feedback control is developed for the model.

Kang and Gershwin [21] study an important cause of inventory record inaccuracy known as stock loss or shrinkage. When stock loss is undetected (*e.g.*, due to theft), it is clear that recorded inventory will overestimate available inventory. The paper considers continuous review $(Q, R)$ inventory policies, and shows via simulation that stockout likelihoods grow substantially in the presence of undetected stock loss; *e.g.*, an increase of stock loss from zero to one percent leads to an increase in stockout likelihood from 0.5% to 15%, and the lost sales percentage due to stock loss can be substantially higher than the stock loss percentage itself. Furthermore, the paper shows that the harmful effects of stock loss are greater in

systems that have short lead times and small order quantities, such as those found in lean supply chain environments. To manage inventory record inaccuracies, the paper proposes several compensation strategies and assesses the benefit of each via comparative simulation.

Uckun at el. [53] consider a single period replenishment problem in a two level supply chain consisting of a retailer and a supplier with multiple warehouses. This research investigates investing to decrease inventory record inaccuracy, under the assumption that investing in technology eliminates inaccuracy. The decision problem then is to determine how many warehouses to outfit with technology, with the objective of maximizing total profit. The primary conclusions obtained from their analysis are that technology investment is much more significant when warehouses do not share inventory, and that investment levels should decrease as demand variance increases.

Kok and Shang [23] propose a model that considers both inventory replenishment and audit costs. The objective is to find a joint replenishment and audit policy that minimizes total cost in a finite horizon. For single period problems, the paper shows the existence of an optimal threshold policy for the audit decision, and a base-stock policy for replenishment. Multiple period problems are addressed using a revised dynamic programming formulation with a cost approximation, and a near-optimal heuristic based on this approximation is proposed. Numerical results indicate that effective policies can substantially reduce the costs of inventory inaccuracies.

Finally, in DeHoratios at el. [12], a Bayesian approach is proposed to address the multiperiod inventory management problem of a single item with uncertain inventory levels. The paper shows that the probability distribution on inventory levels is a sufficient statistic for a dynamic programming formulation of the problem. Using simulations, the paper shows that heuristic Bayesian inventory ordering policies perform well, and that better audit-triggering policies than the traditional zero-walk can be found.

This chapter is outlined as follows. In Sections 2.2 and 2.3 we define the POMDP and present preliminary results. We then present a class of policies for the POMDP, the zero-memory policies, in Section 2.4. A zero-memory policy selects actions on the basis of the most recent state observation. For a zero-memory policy, the expected total discounted

reward to be accrued over the infinite horizon, $v$, is shown to be the inner product of the current state probability mass vector $x$ and a vector $\gamma$ , *i.e.*, $v = x\gamma$ , where $\gamma$ only depends on the most recent state observation.

We then show in Section 2.5 that $\gamma$ can be represented as a power series in $\epsilon$, where $\epsilon$ is the probability of a state observation error. Thus, for small $\epsilon$, $v$ can be approximated by $x(\alpha^0 + \epsilon\alpha^1)$, which implies that whether or not improved state observation quality improves system performance is dependent on the signs of the scalar elements of the vector $\alpha^1$. We show that if $\gamma$ is generated by an optimal policy for the completely observed counterpart (COC) of the POMDP, then all scalar elements of $\alpha^1$ are non-positive; therefore, decreasing $\epsilon$, *i.e.*, improving state observation quality, does not decrease $v$, *i.e.*, does not degrade system performance. Further, we show that if the zero-memory policy does not depend on the most recent state observations , then $\alpha^1 = 0$; hence, system performance is independent of state observation quality for observation invariant policies.

We show by example in Section 2.6 that a near-optimal COC zero-memory policy can produce a vector $\alpha^1$ with all positive scalar elements. Thus, it is possible that what appears to be a good sub-optimal design can degrade system performance, given improved state observation quality. A second example shows that it is possible for the vector $\alpha^1$ to have both positive and negative scalar elements, indicating that the impact of improved state observation quality on system performance may depend on $x$.

In Section 2.7 we consider a single item inventory problem in which inventory observations are assumed to be noise corrupted. We provide the POMDP formulation and its COC. Further, we develop a simulation model to assess the potential benefit of improving accuracy of inventory counts when zero-memory policies are implemented. We conclude this section with a computational analysis.

## *2.2 Problem Definition*

Let $\{s(t), t = 0, 1, \cdots\}$, $\{z(t), t = 1, 2, \cdots\}$, and $\{a(t), t = 0, 1, \cdots\}$ be the state, observation, and action processes respectively. Assume that the state space $S$, the observation space $Z$, and the action space $A$ are each finite and that these three processes are related by the

given probabilities

$$p_{ij}(z,a) = P[z(t+1) = z, s(t+1) = s | s(t) = i, a(t) = a],$$

where $P(z,a) = \{p_{ij}(z,a)\}$, a sub-stochastic matrix which is such that $\sum_z P(z,a)$ is stochastic.

We assume the problem horizon is countably infinite and that for decision epoch $t \in \{0, 1, \cdots\}$, action $a(t)$ can be selected based on $h(t) = \{z(t), \cdots, z(1), a(t-1), \cdots, a(0), x(0)\}$, where $x(t) = \{x_i(t)\}$ is a probability mass vector (*pmv*) and where $x_i(t) = P[s(t) = i | h(t)]$. Note, $x(0)$ is an *a priori pmv*. A policy at decision epoch $t$ is a function $\delta_t : \{d(t)\} \rightarrow A$; a strategy is an ordered sequence of policies $\pi = \{\delta_t, t = 0, 1, \cdots\}$.

Let $r(i,a)$ be the reward accrued at decision epoch $t$, given $s(t) = i$ and $a(t) = a$. The criterion is

$$E_{x(0)} = \left\{ \sum_{t=0}^{\infty} \beta^t r[s(t), a(t)] \right\},$$

where $E_x$ is the expectation operator, conditioned on *pmv* x. In order to insure that the criterion is well defined, assume throughout that $\beta < 1$ and there is an $M$ such that $|r(i,a)| \leq M$ for all $i$ and $a$ (Puterman [40]). The problem objective is to determine a strategy that maximizes the criterion.

### *2.3 Preliminary Results*

It is well-known that $\{x(t), t = 0, 1, \cdots\}$ is a sufficient statistic for the POMDP (Astrom [3]). Let $X = \{x \geq 0 : \sum_{i \in S} x_i = 1\}$, $\| \cdot \|$ be the supremum norm on $X$, and $V$ be the set of all bounded, real-valued functions on $X$. Define $H_\delta : V \rightarrow V$ and $H : V \rightarrow V$ as follows:

$$\begin{aligned} [H_\delta v](x) &= xr[\delta(x)] + \beta \sum_z \sigma(z, x, \delta(x)) v[\lambda(z, x, \delta(x)], \\ [Hv](x) &= \max_{a \in A} \left\{ xr(a) + \beta \sum_z \sigma(z, x, a) v[\lambda(z, x, a)] \right\}, \end{aligned}$$

where $y\mathbf{1} = \sum_i y_i$, $\sigma(z, x, a) = xP(z, a)\mathbf{1}$, and $\lambda(z, x, a) = xP(z, a)/\sigma(z, x, a)$ when $\sigma(z, x, a) \neq$

0. We remark that $x(t+1) = \lambda[z(t+1), x(t), a(t)]$ and $\sigma(z, x, a) = P(z(t+1) = z | x(t) = x, a(t) = a)$. We remark that $Hv = \sup_\delta H_\delta v$.

It is shown in (Puterman [40]) and elsewhere that $H$ and $H_\delta$ are contraction operators on the Banach space $(V, \|\cdot\|)$. Thus, there exists a unique $v^* \in V$ such that $v^* = Hv^*$ and $\lim_{n\to\infty} \|v^n - v^*\| = 0$ for $v^{n+1} = Hv^n$, for $v^0 \in V$. (An analogous statement can be made about $v^*_\delta$ for each $\delta$). Further, if $H_\delta v^* = Hv^*$, then the (stationary) strategy $\pi = \{\delta, \delta, \cdots\}$ is an optimal strategy.

### *2.4 Zero-Memory Policy*

We define a zero-memory policy at decision epoch $t$ as $\delta_t$ composed of decision rules of the form $\delta_t : Z \to A$ such that $a(t) = \delta_t(z(t))$. That is at each decision epoch the decision rule is based on the most current state observation. We remark that $(z(t), a(t-1), x(t-1))$ is a sufficient statistic.

For stationary zero-memory strategy $\pi = \{\delta, \delta \cdots\}$, we note (with a slight abuse of notation) that

$$[H_\delta v](z, a', x') = \overline{x} r(\delta(z)) + \beta \sum_k \sigma(k, \overline{x}, \delta(z)) v[k, \delta(z), \overline{x}],$$

where $x' = x(t-1)$, $\overline{x} = x(t) = \lambda(z, x', a')$, $z = z(t)$, and $a' = a(t-1)$. A simple induction argument, based on the fact that $\lim_{n\to\infty} \|v^n_\delta - v^*_\delta\| = 0$, implies the following preliminary result.

**Lemma 1** *There exists a vector $\gamma$ such that $v^*_\delta(z, a', x') = \lambda(z, x', a')\gamma(z)$, where $\gamma$ is the (unique) solution of*

$$\gamma(z) = r(\delta(z)) + \beta \sum_k P(k, \delta(z))\gamma(k)$$

*for stationary, zero-memory policy $\delta$.*

### *2.5 Policy Performance and Observation Quality*

We now investigate the impact of state observation quality on zero-memory policy performance. Henceforth, assume $S = Z$, the observation probability $q(z|j, i, a) = P[z(t+1) =$

$z|s(t+1)=j, s(t)=i, a(t)=a]$ is independent of $i$ and $a$ and

$$q(z|j) = \begin{cases} 1-\epsilon & \text{if } z=j \\ \sigma_{jz}\epsilon & \text{if } z \neq j \end{cases},$$

where $\sigma_{jz} \geq 0$, $\sigma_{jj}=0$, and $\sum_z \sigma iz = 1$ for all $j$, and $\epsilon > 0$ represents the probability of an inaccurate state observation.

We remark that the stochastic matrix $\{q(z|j)\}$ represents better state observation quality if it depends on $\epsilon$ rather than $\epsilon'$ and $\epsilon < \epsilon'$. See (White and Harrington [55]) for another closely related description of state observation quality.

Note that $p_{ij}(z,a) = q(z|j)p_{ij}(a)$, where $p_{ij}(a) = \sum_z p_{ij}(z,a) = P[s(t+1)=j|s(t)=i, a(t)=a]$ which is often referred to as the transition probability.

Intuitively, we would expect $v^*_\delta(z,a',x')$ to increase (or at least not decrease) as $\epsilon$ gets smaller. We show below that this characteristic is not in general true but present conditions that guarantee it holds. We now show that for small $\epsilon$, the expected total discounted reward for a zero-memory stationary strategy can be represented by a power series in $\epsilon$.

**Proposition 1** *Assume* $\epsilon < \frac{(1-\beta)}{2\beta}$. *Then,* $\gamma(z) = \sum_{l=0}^{\infty} \epsilon^l \alpha^l(z)$, *where:*

$$\begin{aligned} \alpha^0(i,z) &= r(i,\delta(z)) + \beta \sum_j p_{ij}(\delta(z))\alpha^0(j,j), \\ \Delta^l(i,z) &= \beta \sum_j p_{ij}(\delta(z)) \left[ \sum_{z \neq j} \sigma_{jz}\alpha^l(j,z) - \alpha^l(j,j) \right], \quad l \geq 0 \\ \alpha^l(i,z) &= \Delta^{l-1}(i,z) + \beta \sum_j p_{ij}(\delta(z))\alpha^l(j,j), \quad l \geq 1. \end{aligned}$$

**Proof:** By successive approximations. It follows from Lemma 1 that $lim_{n\to\infty}\|\gamma_n(z) - \gamma(z)\| = 0$, where

$$\gamma_{n+1}(z) = r(\delta(z)) + \beta \sum_k P(k,\delta(k))\gamma_n(k),$$

and where $\gamma_0 = 0$. It follows from the definition of $q(z|j)$ that

$$\gamma_{n+1}(i,z) = r(i,\delta(z)) + \beta \sum_j p_{ij}(\delta(z))\gamma_n(j,j) + \epsilon\Delta_n(i,z),$$

where

$$\Delta_n(i,z) = \beta \sum_j p_{ij}(\delta(z)) \left[ \sum_{k \neq j} \sigma_{jk}\gamma_n(j,k) - \gamma_n(j,j) \right].$$

It is then straightforward to show that $\gamma_n(i,z) = \sum_{l=0}^{n} \epsilon^l \alpha_n^l(i,z)$, where:

$$\begin{aligned} \alpha_{n+1}^0(i,z) &= r(i,\delta(z)) + \beta \sum_j p_{ij}(\delta(z))\alpha^0(j,j) \\ \alpha_{n+1}^l(i,z) &= \Delta_n^{l-1}(i,z) + \beta \sum_j p_{ij}(\delta(z))\alpha_n^l(j,j) \end{aligned}$$

for $l = 1, \cdots, n$, where $\alpha_{n+1}^{n+1}(i,z) = \Delta_n^n(i,z)$, and for $l = 0, \cdots, n$,

$$\Delta_n^l(i,z) = \beta \sum_j p_{ij}(\delta(z)) \left[ \sum_{k \neq j} \sigma_{jk}\alpha_n^l(j,k) - \alpha_n^l(j,j) \right].$$

Letting $n \to \infty$ gives the result, assuming $\lim_{n\to\infty} \sum_{l=0}^{n} \epsilon^l \alpha_n^l(i,z)$ exists.

We now show that the infinite sum is well-defined if $\epsilon < \frac{(1-\beta)}{2\beta}$. Since $|r(i,a)| \leq M$ for all $i$ and $a$, it follows that $\|\alpha^0\| \leq \frac{M}{1-\beta}$, where $\|\cdot\|$ is the supremum norm. A similar argument implies that for $l \geq 1$, $\|\alpha^l\| \leq \frac{\|\Delta^{l-1}\|}{1-\beta}$, where we note that $\|\Delta^{l-1}\| \leq 2\beta\|\alpha^{l-1}\|$. Thus, $\|\alpha^l\| \leq [\frac{2\beta}{1-\beta}]^l(\frac{M}{1-\beta})$. Convergence of the infinite series is then guaranteed if $\frac{2\beta\epsilon}{1-\beta} < 1$.

■

We note that determining $\alpha^0$ is computationally identical to computing the expected total discounted reward to be accrued over the infinite horizon generated by $\delta$ for the completely observed (*i.e.*, COC) case. Computing $\Delta^{l-1}$ and $\alpha^l$, for each $l \geq 1$, has substantially more modest computational requirements.

Our numerical results thus far indicate that $\|\alpha^l\|$ approaches zero quickly as $l$ grows large. Thus, $\frac{1-\beta}{2\beta}$ appears to be a very conservative upper bound on $\epsilon$ in order to guarantee that the limit $\lim_{n\to\infty} \sum_{l=0}^{n} \epsilon^l \alpha^l$ exists. Further, for small $\epsilon$ (*i.e.*, reasonably accurate observation quality), $\alpha^0 + \epsilon\alpha^1$ appears to be a good approximation of $\gamma$. Recall by Lemma 1 that the

expected total discounted reward over the infinite horizon, $v$, is such that $v = \lambda\gamma$, where $\lambda$ is a *pmv*. Hence, if $\alpha^1 \leq 0 (\alpha^1 \geq 0)$, then $v$ is non-decreasing (non-increasing) as $\epsilon$ gets smaller. Thus, if the finite-memory policy is such that $\alpha^1 \leq 0$, then there may be value in improving state observation quality in order to improve expected system performance. However, if $\alpha^1 \geq 0$, then improving state observation quality will never improve, and may degrade expected system performance. If $\alpha^1$ is neither non-positive nor non-negative, then whether there is value in improving state observation quality depends on the sign of $\lambda\alpha^1$ and hence the value of $\lambda$.

We now present a zero-memory policy that guarantees $\alpha^1 \leq 0$. This policy is identical to the optimal policy for the case where $\epsilon = 0$ (*i.e.*, the perfect state observation case and hence is a COC), and is thus relatively easy to calculate and implement.

**Corollary 1** *Let $\delta^* : Z \rightarrow A$ be a policy that achieves the maximum in*

$$\alpha^*(i) = \max_{a \in A} \left\{ r(i,a) + \beta \sum_j p_{ij}(a)\alpha^*(j) \right\}.$$

*Then, $\alpha^*(i) = \alpha^0(i,i) \geq \alpha^0(i,z)$, for all $i$ and $z$, and hence $\alpha^1(i,z) \leq 0$ for all $i$ and $z$.*

**Proof:** Let $\{\alpha_n^*\}$ and $\{\delta_n^*\}$ be such that

$$\begin{aligned} \alpha_{n+1}^*(i) &= \max_{a \in A} \left\{ r(i,a) + \beta \sum_j p_{ij}(a)\alpha_n^*(j) \right\} \\ &= r(i,\delta_n^*(i)) + \beta \sum_j p_{ij}(\delta_n^*(i))\alpha_n^*(j), \end{aligned}$$

where $\alpha_0^*(i) = 0$. Then, $\lim_{n\rightarrow\infty} \|\alpha_n^* - \alpha^*\| = 0$ and hence $\alpha^*(i) = \alpha^0(i,i)$. It then follows from Proposition 1 that for the case where $z(t) = z \neq i$, that

$$\begin{aligned} \alpha^0(i,z) &= r(i,\delta^*(z)) + \beta \sum_j p_{ij}(\delta^*(z))\alpha^*(j) \\ &\leq r(i,\delta^*(i)) + \beta \sum_j p_{ij}(\delta^*(i))\alpha^*(j) = \alpha^*(i). \end{aligned}$$

Clearly, $\alpha^0(i,z) = \alpha^*(i,z)$. The fact that $\alpha^1(i,z) \leq 0$ follows from the definition of $\alpha^1$ in terms of $\Delta^0$ and hence in terms of $\alpha^0$.

■

We remark that we showed in the proof of Corollary 1 that $\alpha^*(i)$ is an upper bound on $\alpha^0(i,z)$ for any $z$, where $\alpha^0$ is generated using $\delta^*$. More generally, $\alpha^*$ is an upper bound on $\alpha^0$ for any $z$ and any zero-memory policy, which is true due to the fact that for the COC decision process (where $z(t) = s(t)$), $s(t)$ is a sufficient statistic for $z(t)$. We also remark that if we select $\delta^*$ to achieve the minimum, rather than the maximum, in the optimality equation in Corollary 1, then $\alpha^1(i,z) \geq 0$ for all $i$ and $z$. Thus, as we are assured that there exists a zero-memory policy whose performance will not degrade as state observation quality improves, we are also assured that there exists a zero-memory policy whose performance will not improve as state observation quality improves.

We now examine a class of policies, history invariant policies, and show that such policies are independent of state observation quality. Thus, improving or degrading state observation quality will have no effect on the performance of a stationary history invariant policy.

**Corollary 2** *Assume $\delta$ is a such that $\delta(z) = a$ for all $z$. Then $\alpha^l = 0$ for all $l \geq 1$ and hence the expected total discounted reward is independent of $\epsilon$.*

**Proof:** If $\delta(z) = a$ for all $z$, then $\alpha^0 = (I - \beta P(a))^{-1} r(a)$, where $P(a) = \{p_{ij}(a)\}$ and $r(a) = \{r(i,a)\}$. Thus, $\alpha^0$ is independent of $z$, which implies $\Delta^0 = 0$ and hence $\alpha^1 = 0$. An induction argument then implies that $\Delta^{l-1} = 0$ and hence $\alpha^l = 0$ for all $l \geq 1$.

■

## *2.6 Examples*

We now present two inventory control examples that only differ in the observation probability distribution. Let the per unit per period holding cost $h = 1999$, the per unit ordering cost $c = 1000$, the per unit selling price $p = 3000$ and the discount factor $\beta = 0.9$. Thus $r(i,a) = -hi - ca + p[\sum_{j \leq i+a} jP(j) + \sum_{j>i+a}(i+a)P(j)]$, where $P(j)$ is the demand probability distribution and is given by:

$$P(j) = \begin{cases} 0.5 & \text{if } j = 1 \\ 0.45 & \text{if } j = 2 \\ 0.05 & \text{if } j = 10 \\ 0 & \text{otherwise} \end{cases}$$

For the first example, the observation probability matrix is

$$q(z|j) = \begin{cases} 1-\epsilon & \text{if } z = j \\ \frac{\epsilon}{3|j-z|} & \text{if } 2 \leq j \leq 8, \quad z \neq j, \quad j-2 \leq z \leq j+2 \\ \frac{\epsilon}{3|j-z|} & \text{if } j = 1, \quad 2 \leq z \leq 3 \\ \frac{1}{2}\epsilon & \text{if } j = 1, \quad z = 0 \\ \frac{2\epsilon}{3|j-z|} & \text{if } j = 0, \quad 1 \leq z \leq 2 \\ \frac{\epsilon}{3(j-z)} & \text{if } j = 9, \quad 7 \leq z \leq 8 \\ \frac{1}{2}\epsilon & \text{if } j = 9, \quad z = 10 \\ \frac{2\epsilon}{3(j-z)} & \text{if } j = 10, \quad 8 \leq z \leq 9 \\ 0 & \text{otherwise} \end{cases}$$

We now examine the behavior of zero-memory policies of the following form:

$$\delta(z) = \begin{cases} k-z & \text{if } z \leq k \\ 0 & \text{otherwise} \end{cases}.$$

Numerical calculations imply that $\alpha^1(i,j) \leq 0$ for all $i,j$ for $k \in \{0,1,2\}$, implying that these policies improve system performance as state observation quality increases for sufficiently small $\epsilon$. We also observe that $\alpha^1(i,j) \geq 0$ for all $i,j$ for $k \in \{9,10\}$, implying that these policies decrease system performance as state observation accuracy increases. For $k \in \{3,4,5,6,7,8\}$ we observe $\alpha^1(i,j) \leq 0$ for some $i,j$ and we also observe $\alpha^1(i,j) \geq 0$ for other $i$ and $j$ implying that the impact of these policies, given improved state observation accuracy, is inconclusive. We remark that the optimal completely observed MDP order-up-to level is 2, which is consistent with Corollary 1. We also note that $\alpha^1(i,j) = 0$ for all $i,j$ for $k = 0$ which is consistent with Corollary 2.

We now consider the same parameter values except that the observation probability matrix is given by:

$$q(z|j) = \begin{cases} 1-\epsilon & \text{if } z = j \\ \frac{\epsilon}{10} & \text{otherwise} \end{cases}$$

For this example we examine policies of the same form described above. Numerical calculations imply that $\alpha^1(i,j) \leq 0$ for all $i,j$ for all $k \in \{0,1,2\}$, implying that the system performance of these policies improves with improved observation quality for sufficiently small $\epsilon$. We also observe that $\alpha^1(i,j) \geq 0$ for all $i,j$ for all $k \geq 3$, implying that these policies will degrade system performance if given improved state observation quality. We remark that the optimal completely observed order-up-to level is again 2. We also note that $\alpha^1(i,j) = 0$ for all $i,j$ for $k = 0$.

We recall from Corollary 1, the "order-up-to 2" policy is guaranteed to improve systems performance if given improved state observation quality. Since state observations may be inaccurate, it would seem reasonable to use an "order-up-to $Y$" policy for some $Y > 2$. Interestingly, our numerical results indicate that if we did so, then the resulting policy would be guaranteed not to improve systems performance if given improved state observation quality.

## *2.7 Computational Study of Zero-Memory Policies for Inventory Systems*

We have already shown that a zero-memory policy for the POMDP that is equivalent to an optimal policy for the COC is adaptive (*i.e.*, system performance will not degrade as the observation quality improves) when the probability of the observation error is sufficiently small. The main objective of this section is to quantify the potential benefit of improving the accuracy of state observability when this type of zero-memory policy is employed in the context of inventory management. Specifically, we consider a single item periodic review inventory system with inaccurate inventory counts, stochastic demand, and lost sales.

Inventory systems with perfectly accurate counts face only variability associated with the demand process. In the presence of inaccurate inventory counts there is another source

of variability due to this inaccuracy. There are two alternative approaches to hedge against the added variability due to the imperfect inventory counts: (1) invest in improving the accuracy of the counts and (2) increasing stocking levels. We will investigate two different types of zero-memory policies that capture these two alternatives.

### 2.7.1 Model Formulation

Consider a single product periodic review inventory system in which a decision maker selects a replenishment quantity at each decision epoch in order to maximize expected total discounted profit over an infinite planning horizon. At the beginning of each decision epoch, an observation of the current inventory level becomes available to the decision maker; this observation may be noise corrupted. Selection of the replenishment quantity at the current epoch is based on all past and present inventory observations and all past ordering decisions. We assume that the quantity ordered is received immediately (no leadtime), no backlogging is permitted and that demand in any given period is stochastic, independent and identically distributed with known probability distribution.

More precisely, let $s(t)$ be the inventory level at decision epoch $t$ just prior to the selection of the replenishment decision $a(t)$. Let $d(t)$ be the demand realized between time $t-1$ and time $t$ and let $P(i) = P(d(t) = i)$ for $i \in \{0, \cdots, D\}$ and for all $t$. We assume replenishment decisions are made for each $t \in \{0, 1, \cdots\}$. Let $z(t)$ be the partially noise corrupted observation of the inventory level that is available to the decision maker just prior to the selection of $a(t)$ and let $h(t) = \{z(t), \cdots z(1), a(t-1), \cdots, a(0), x(0)\}$, where $x(0) = \{x_i(0)\}$ and $x_i(0) = P(s(0) = i)$. Thus, $x(0) \in \mathbb{X} = \left\{x \geq 0 : \sum_{i=0}^{D} x_i = 1\right\}$ is the *a priori* probability distribution of $s(0)$. Note that $h(t)$ represents all the information available to the decision maker prior to the selection of $a(t)$.

The observation probabilities are assumed to be dependent of parameters $k$ and $\epsilon$ as follows:

$$P^k(z|j) = \begin{cases} \frac{\epsilon}{\min\{D,j+k\}-(j-k)^+} & z \neq j, (j-k)^+ \leq z \leq \min\{D, j+k\} \\ 1-\epsilon & z = j \\ 0 & otherwise \end{cases}, \quad (1)$$

where $\epsilon$ is the probability of an inaccurate observation and $k$ the maximum possible difference between the observed and the true inventory level.

Let $p$, $c$ and $h$ be the per unit selling price, per unit ordering cost and per unit per period holding cost respectively. We assume that the holding cost from time $t$ to $t+1$ is given by $hs(t)$.

The *Inaccurate Inventory Replenishment Problem* (*IIRP*) is to find a strategy that maximizes the following criterion:

$$\mathbf{E}^{\pi}_{x(0)} = \left\{ \sum_{t=0}^{\infty} \beta^t r[s(t), a(t)] \right\},$$

where $\mathbf{E}^{\pi}_{x(0)}$ is the expectation operator conditioned on $x(0)$ and use of strategy $\pi$. Parameter $\beta$ represents the discount factor and $r[s(t), a(t)] = -hs(t) - ca(t) + p\mathbf{E}\Big\{ \min\{d(t+1), s(t)\} \Big\}$.

The optimality equation is then:

$$v(x) = \max_a \left\{ \sum_i x_i r(i,a) + \beta \sum_z \sigma(z,x,a) v(\lambda(z,x,a)) \right\},$$

where

$$\sigma(z,x,a) = \sum_i x_i \sum_j P(z|j) P(j|i,a) \neq 0,$$

$$\lambda(z,x,a) = \left\{ \lambda_j(z,x,a) = \frac{\sum_i x_i P(z|j) P(j|i,a)}{\sigma(z,x,a)} \right\},$$

and

$$P(j|i,a) = \begin{cases} 0 & \text{if} \quad j > i+a \\ P(i+a-j) & \text{if} \quad 1 \leq j \leq i+a \\ \sum_{k \geq i+a} P(k) & \text{if} \quad j = 0 \end{cases}.$$

### 2.7.2 Two Zero-Memory Policies

We investigate the performance of two classes of zero-memory policies, (i) the optimal policy for the COC problem denoted by (COC$^*$), and (ii) a policy that we refer to as COC overstocking zero-memory policy. The COC$^*$ policy for the *IIRP* is given by the following optimality equation:

$$v^C(i) = \max_a \left\{ r(i,a) + \beta \sum_j P(j) v^C([i+a-j]^+) \right\}$$

We remark that an optimal policy COC$^*$ is an order-up-to policy. That is, there exists an integer $y$ such that:

$$a(i) = \begin{cases} y - i & \text{if } i \leq y \\ 0 & \text{otherwise} \end{cases}.$$

We define the COC overstocking zero-memory policy denoted by COC$^o(\gamma)$ as follows: Let $\pi = \{\delta\}$ be the COC$^*$ zero-memory policy where $a = \delta(z)$, then $\pi^o(\gamma) = \{\delta^o_\gamma\}$ is the COC$^o(\gamma)$ zero-memory policy where $a = \delta^o_\gamma(z) = \delta(z) + \gamma$ and $\gamma > 0$.

### 2.7.3 Numerical Experiments

In this section, we present the results of numerical experimentation using both the COC$^*$ and the COC$^o(\gamma)$ policies for the *IIRP*. Specifically, we want to understand the impact of three problem features on the value of improved inventory accuracy: (1) demand variability measured as the coefficient of variation of the demand process, (2) the relative contribution of holding cost to overall cost and (3) variability of the observation process measured as the maximum absolute error of the observation quantity which is represented by parameter $k$ (see Equation 1).

We use the value iteration method to solve the COC of the *IIRP* problem. In order to evaluate system performance for any zero-memory policy we employ steady state simulation. Based on initial pilot runs, it was estimated that these systems need 1250 periods of simulation warm-up to reach steady state. Therefore, in each replication a total of 2500 periods are simulated, and statistics are generated only over the final 1250 periods. Results

from the pilot runs suggest that a total of 30 replications are required to reach a relative error of 0.01 for a confidence level of 99% (see Law and Kelton [25], Chapter 9).

Scenarios are generated with the per unit ordering cost and per unit selling price set to be $c = 160$ and $p = 200$. The per period per unit holding cost $h$ is varied among scenarios and is selected from the set $\{1.6, 8, 16\}$. The per period demand $B(n, \rho)$ is assumed to be Binomially distributed with $n = 20$ and $\rho$ selected from the set $\{0.2, 0.5, 0.8\}$. The discount factor is assumed to be $\beta = 0.99$. The levels of inaccuracy $\epsilon$ are selected from the set $\{0.5, 0.4, 0.3, 0.2, 0.1, 0.05, 0.025, 0.0125, 0.00625, 0\}$, and the parameter $k$ used to generate the observation probabilities is selected from the set $\{2, 5, 10, 20\}$. The order-up-to levels of the COC$^*$ policy for each scenario are shown in Table 1. The value of $\gamma$ for the COC$^o(\gamma)$ zero-memory policy is set equal to 2 in all scenarios. We measure the potential value of improve inventory accuracy as the Percentage Reduction in System Profit due to inaccurate counts ($PRSP$), defined as the difference between the profit obtained under perfect accuracy and the profit obtained under inaccuracy, divided by the former value.

**Table 1:** Optimal Order-up-to Levels of COC$^*$

| $\rho$ | 0.2 | 0.2 | 0.2 | 0.5 | 0.5 | 0.5 | 0.8 | 0.8 | 0.8 |
|---|---|---|---|---|---|---|---|---|---|
| $h$ | 1.6 | 8 | 16 | 1.6 | 8 | 16 | 1.6 | 8 | 16 |
| Order-up-to Level | 7 | 6 | 5 | 13 | 12 | 11 | 18 | 18 | 17 |

First, the empirical results match what is predicted theoretically. In all scenarios we observe that the COC$^*$ zero-memory policy is adaptive for small values of $\epsilon$. Furthermore, the computational results indicate that the policy remains adaptive for large values of $\epsilon$. Figure 1 and Figure 2 graphically depict this result; as information quality improves from right to left, the profit loss due to inaccurate counts decreases.

In all instances the ratio $PRSP$ to $\epsilon$ was observed to be always smaller than 1, indicating that an increase in inaccuracy counts of $x\%$ will degrade the system performance by no more than that percentage. We also observed that this ratio increases as the inventory holding increases, indicating that systems with higher inventory holding costs can benefit more from improving accuracy of inventory counts than systems with lower inventory holding costs.

The experiments indicate that demand variability has a significant impact on the benefit

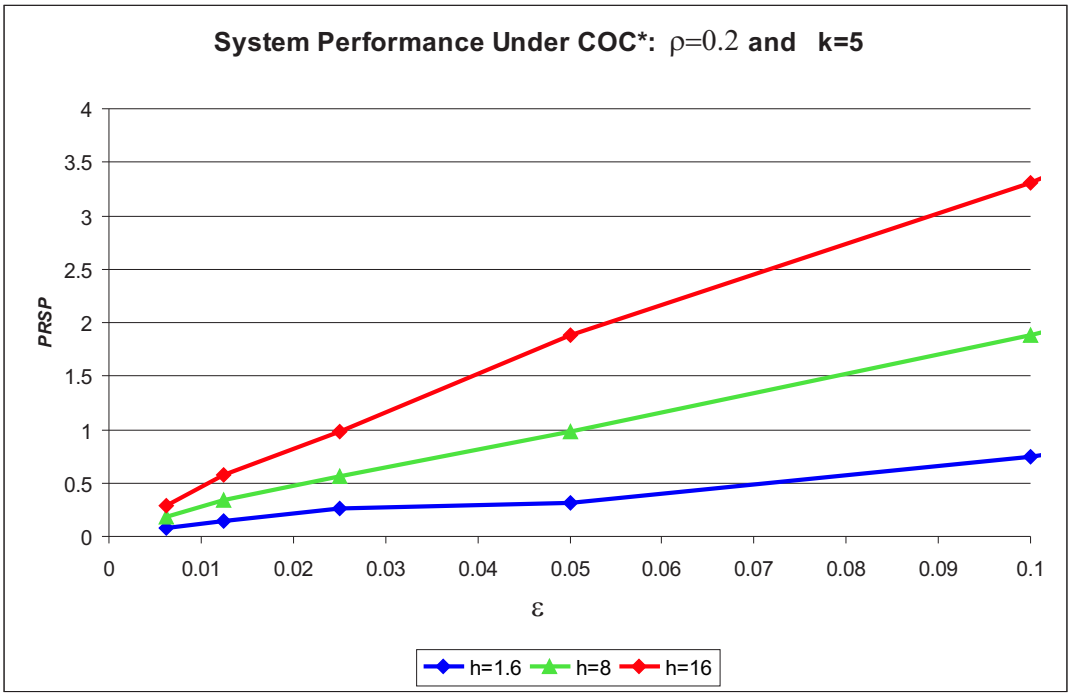



**Figure 1:** Percentage Reduction in Expected Profit Due to Inaccurate Inventory Counts for Small Values of $\epsilon$ and Different Levels Holding Costs ($\rho = 0.2, k = 5$)

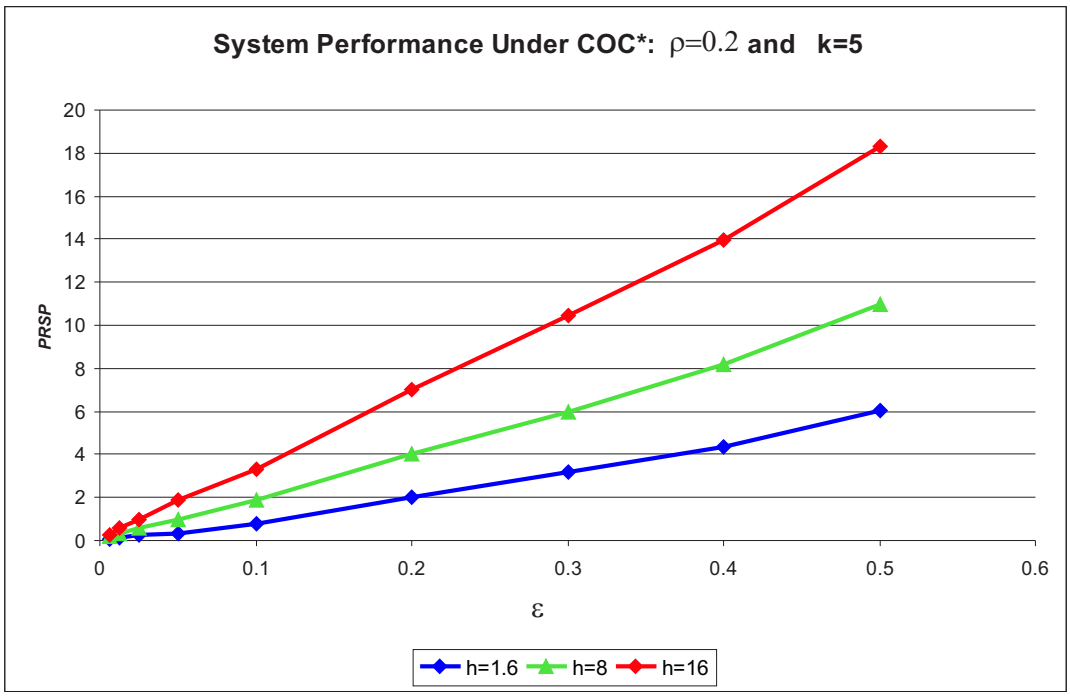



**Figure 2:** Percentage Reduction in Expected Profit Due to Inaccurate Inventory Counts for Different Levels Holding Costs ($\rho = 0.2, k = 5$)

of improving inventory accuracy. For a fixed inventory holding cost $h$ and fixed value of $k$, numerical results show that the benefit of improving inventory accuracy increases as demand variability increases. This behavior is more pronounced as the value of $k$ decreases. This result suggests that as the variability of the observation process ($k$) decreases, the effect that demand variability has on the benefit of improving inventory accuracy is more significant. Figure 3 depicts this result; each graph exhibits the $PRSP$ for the three levels of demand variability, a fixed value of $k$ and $h = 16$. Similar results are observed for other holding cost values.

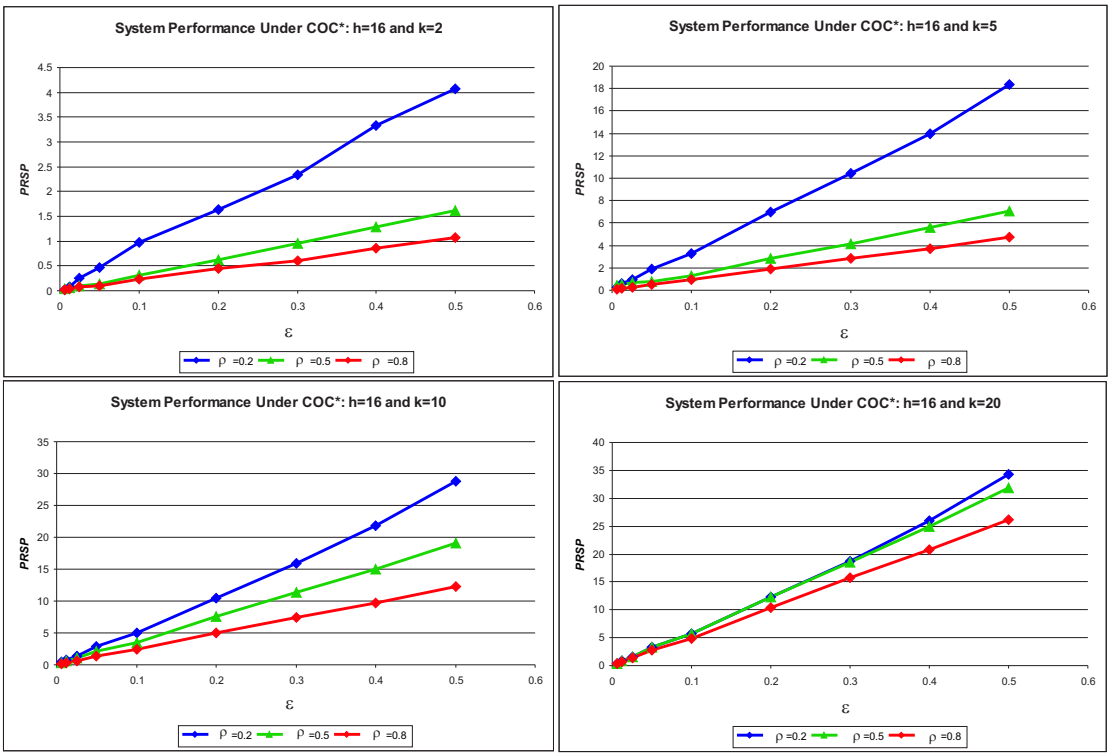



**Figure 3:** Percentage Reduction in Expected Profit Due to Inaccurate Inventory Counts for Different Levels of Demand Variability and Observation Variability ($h = 16$)

Variability in the inaccuracy of the observation process seems to have a significant effect on the benefit of improving inventory counts. As expected, as this variability increases the benefit of improved accuracy also increases (see Figure 4). Furthermore, for the lowest level of inaccurate observation variability ($k = 2$), the COC* zero-memory policy outperformed the COC$^o$(2) zero-memory policy for every level of inaccuracy. However, the percentage

difference in the performance of the two policies for this case is not very significant (see Figure 5).

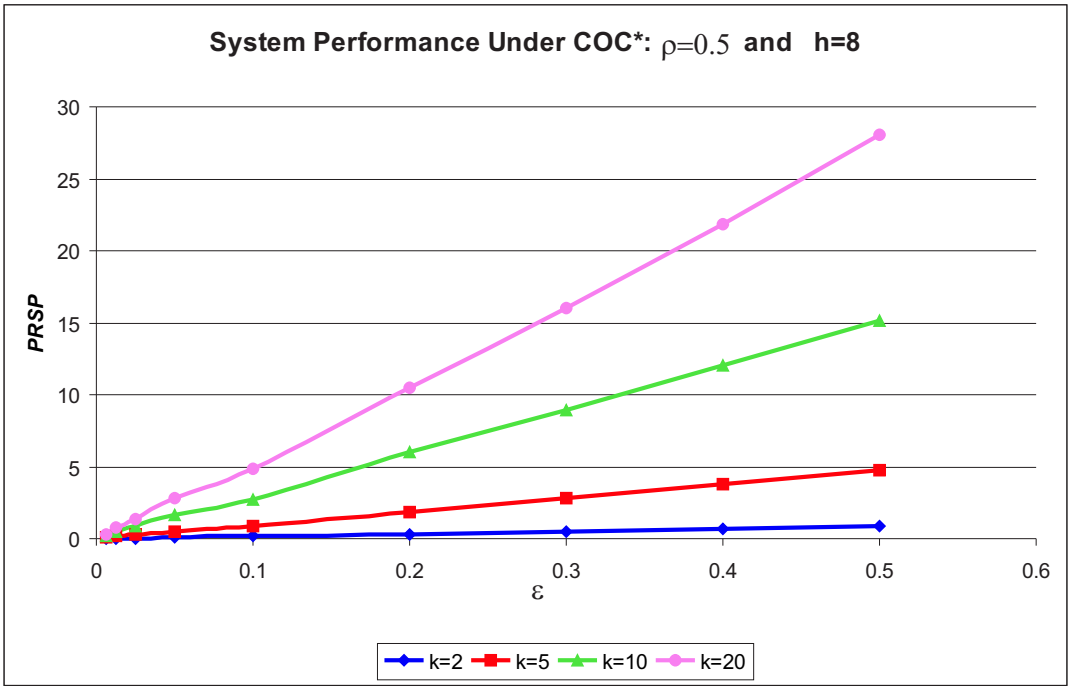



**Figure 4:** Percentage Reduction in Expected Profit Due to Inaccurate Inventory Counts for Different Levels of Observation Variability ($\rho = 0.5, h = 8$)

In contrast, for systems with higher levels of observation variability when $\epsilon \leq 0.10$ either the COC$^*$ zero-memory policy outperformed the COC$^o(2)$ zero-memory policy (in systems with $h \geq 8$) or the relative benefit of the COC$^o(2)$ policy was less than 1% (in systems with h=1.6). In each scenario, we found that there is a break point in the value of $\epsilon$ after which the COC$^o(2)$ starts to outperform the COC$^*$ zero-memory policy; as the value of holding cost increases, the break point value increases (see Figure 6). This result is intuitively clear since there is a tradeoff between the benefit obtained by hedging against the variability due to inaccurate counts by overstocking and the cost of stocking more. Finally, we found that the COC$^o(2)$ zero-memory policy was not always adaptive; Figure 7 clearly demonstrates this fact for the scenarios with parameters $h = 8, k = 2$ and $\rho = 0.8$.

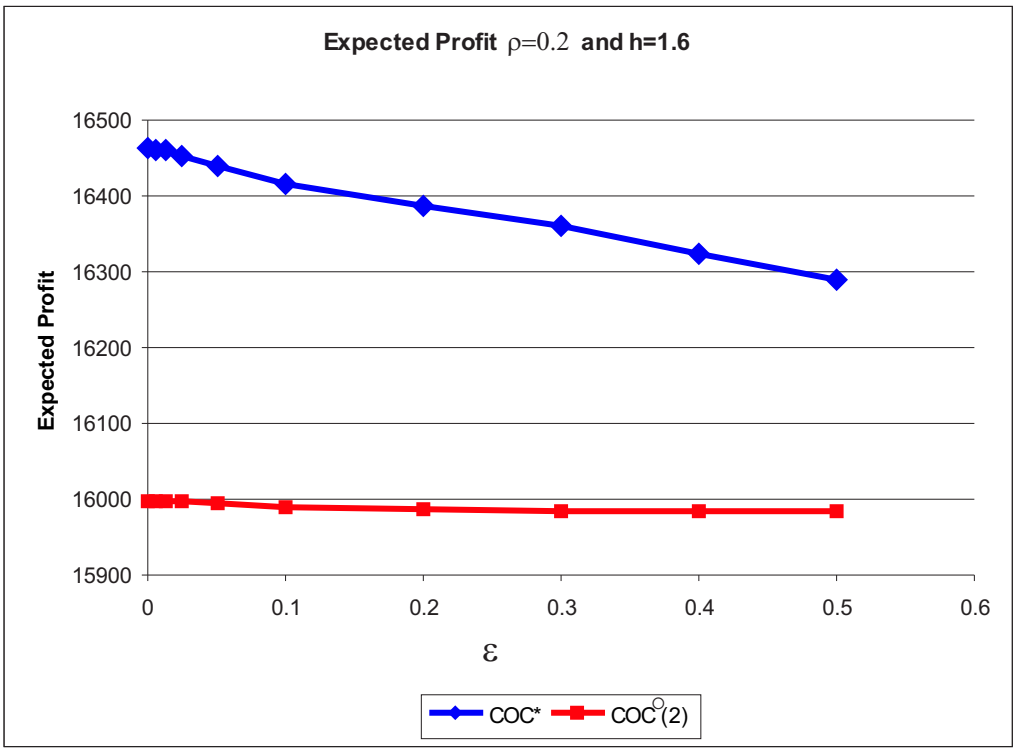
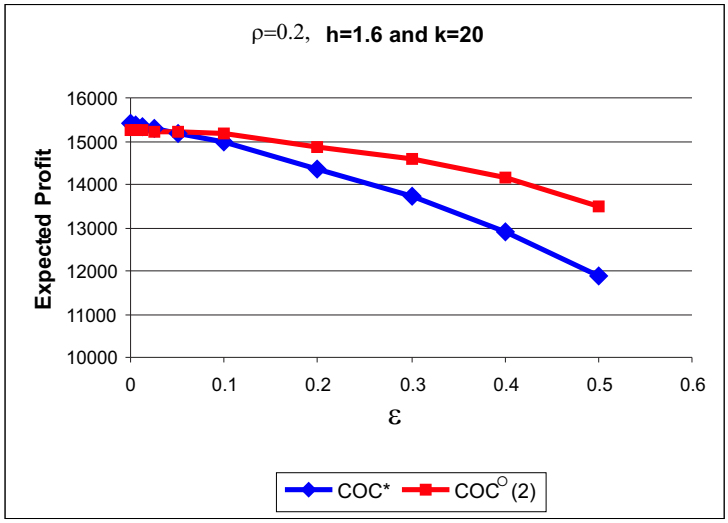
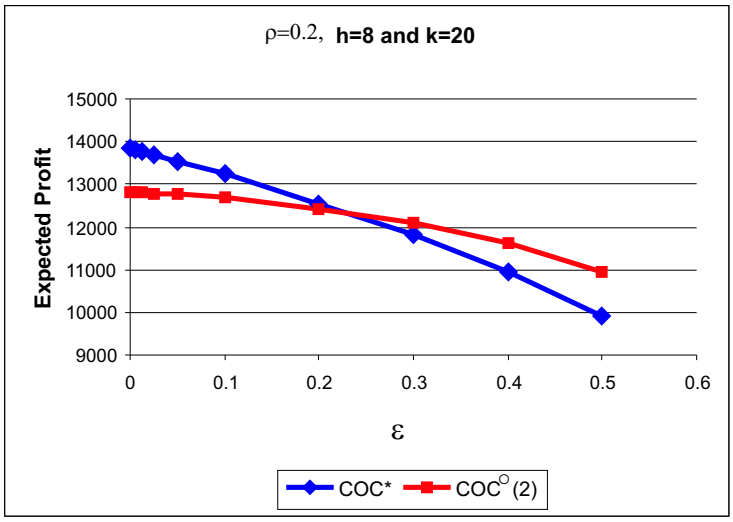



**Figure 5:** Expected Profit Under COC* and COC$^o$(2) Zero-Memory Policy ($h = 1.6, k = 2$ and $\rho = 0.2$



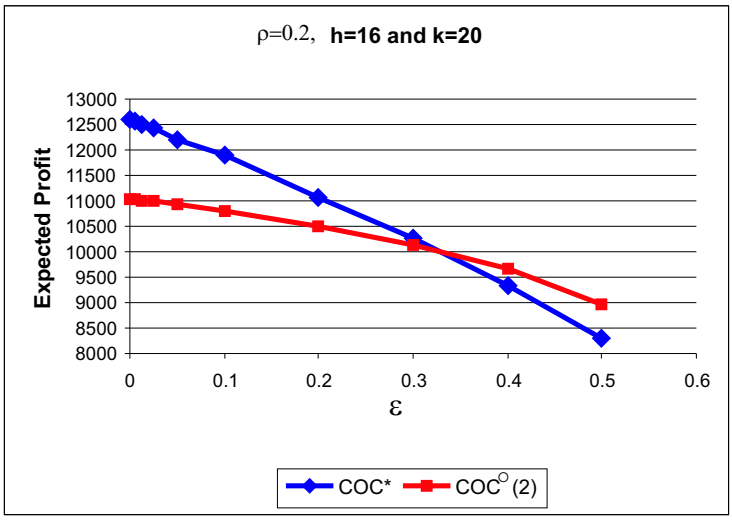




**Figure 6:** Expected Profit for COC* and COC$^o$(2) Zero Memory Policies ($\rho = 0.2$ and $k = 20$)

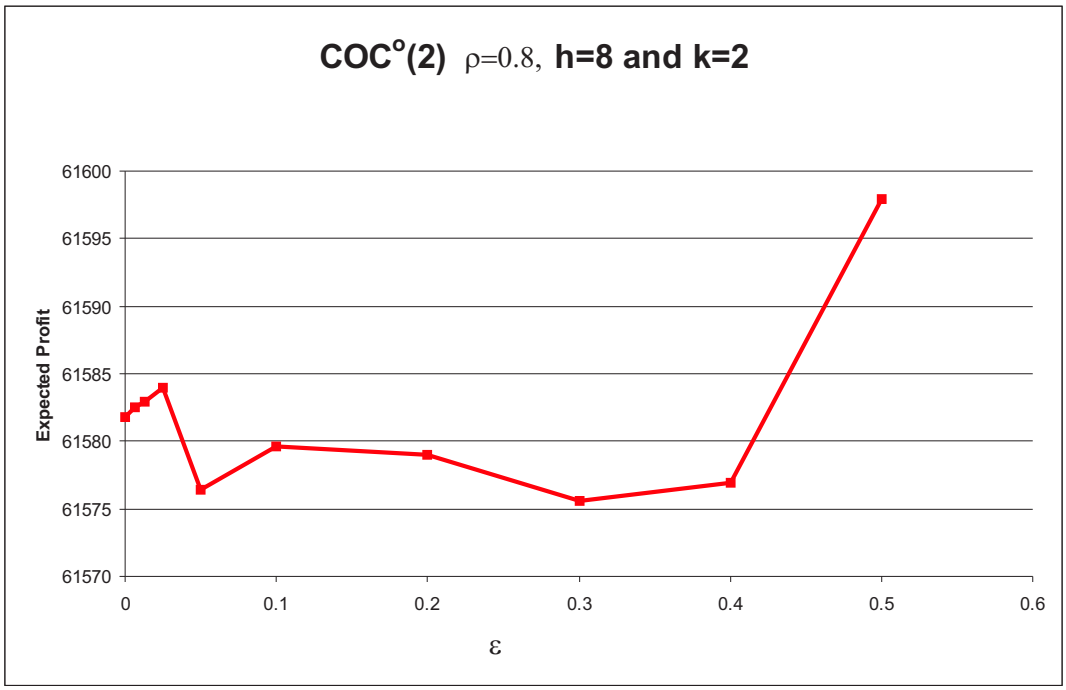



**Figure 7:** Expected Profit for COC$^o$(2) Zero-Memory Policy ($\rho = 0.8, h = 8$ and $k = 2$)

# CHAPTER III

# BOUNDING THE VALUE OF IMPROVING DEMAND OBSERVABILITY FOR A SINGLE ITEM INVENTORY CONTROL WITH MARKOVIAN DEMAND AND LOST SALES

## *3.1 Introduction*

Profitably supplying product to meet customer demand is a crucial objective of virtually all supply chains. Decision models for the sequential control of supply chain systems usually require models of demand and demand observation. In this chapter, we consider periodic inventory control decisions for systems where unsatisfied customer demand is not necessarily completely observed and is lost (*i.e.*, no backlogging). Although true demand may not be observed, we assume that customer sales data, inferred from inventory levels, are perfectly observed.

Perfectly observed sales data may provide censored observations of demand. For a firm selling a single product, if the inventory level remains positive during a period, then sales during that period provide an observation of actual demand. However, if the inventory level drops to zero, sales are a lower bound on actual demand. When firms supply multiple products to customers who may substitute if their first choice is unavailable, the situation is more complex, since sales may represent an artificially inflated observation of demand for a product that customers purchase as a substitute for an unavailable product.

Since sales data alone may provide only partial information about customer demand, many firms develop additional mechanisms aimed at improving understanding of demand. Effective *demand sensing* programs that identify and interpret certain market signals may provide valuable input to the production and inventory control activities of a firm. Examples of such market signals include web hit data for product-specific pages at a firm's internet site or the number of phone inquiries about a specific product to a sales call center. Note that demand sensing data may not only be useful for forecasting future product demand,

but also for improving true demand *observability* during periods with potential lost sales.

Demand sensing programs have a cost, and thus it is important to understand their potential benefit. In this chapter, we focus on understanding the value of improved demand observability for inventory control. To initiate such a study, we consider a very simple supply system for a single product where a single capacity-constrained production site supplies a single consumption site. The consumption site operates a periodic review inventory control system, where orders placed at the end of a period are available at the beginning of the following period. The demand process $d(t)$ for product at the consumption site is non-stationary and discrete, and assumed to be described by a stationary Markov chain with known transition probabilities $P\{d(t+1) = j \mid d(t) = i\}$ (we remark that such probabilities may not be known precisely, and as a result, we treat them as parameters in the numerical analysis presented later). Using a Markov process to model demand is natural when demands may be correlated in time.

For such a supply system, prior period demand is observed completely only when no inventory shortage occurs, and therefore we develop a partially-observed Markov decision process (POMDP) control model. Sales and inventory quantities are assumed to be perfectly observed; thus, this information provides a censored observation of true demand $d(t)$. In addition, we assume that we also receive direct demand observations from a demand sensing program; we summarize the (possibly noise corrupted) observation information provided by demand sensing using the notation $z(t)$.

We first develop a general mathematical model for the inventory control problem with partially-observed Markovian demand and lost sales and present an algorithm for determining an optimal policy. Since this model is intractable for reasonable problem sizes, we next propose three computationally attractive heuristic solution procedures for this problem that produce near-optimal decisions, the third of which is based on a non-standard sufficient statistic with characteristics unique to this problem. We then analyze two extreme cases of the POMDP model, the completely-observed case and the sales-only-observed case, and use them to determine the expected maximum added value of improved demand observability. Finally, we conduct a computational study using these extreme cases in an attempt to

obtain computational evidence regarding the maximum value of demand observability for realistic problem settings.

## *3.2 Related Literature*

Inventory control problems for supply chain systems facing uncertain demands have been studied extensively in the research literature; see Lee and Nahmias [26] for a comprehensive review of single product, single-location inventory models. Demand is usually assumed to be completely observable, even in periods with inventory stockouts; backorders are generated, or lost sales are penalized. Furthermore, most research assumes that demand can be modeled as a stationary and independent process with a known probability distribution. In other cases (for example, when a new product is introduced to the market), researchers assume that this stationary distribution has unknown parameters and use statistical techniques to estimate them; Bayesian approaches for such problems were first proposed and refined by Scarf ([43] and [44]), Karlin [22], Iglehart [19] and later generalized by Azoury [4].

Fewer models consider the case where lost demand is not observed, although many supply chains that serve end consumers share this characteristic. Fisher at el. [15] note that this is a common dilemma faced by retail supply chains, and discuss how ignoring lost sales makes it difficult to set optimal inventory levels, leading to extra current costs and potentially additional lost sales in the future. When lost sales are not observed, the problem of determining the true underlying demand distribution becomes more difficult since the data is censored. Two approaches in the literature use focus on this problem using classical statistical estimation: Nahmias [35] estimates parameters of a normal demand distribution from sales observations assumed to form a right-censored sample, while Agrawal and Smith [2] extends this work to the case of a negative binomial demand distribution.

Other work on problems with unobserved lost sales focuses on the joint problem of distribution parameter estimation and optimal stocking policy determination. Lariviere and Porteus [24] develops a tractable Bayesian approach for updating an *a priori* demand distribution in this setting by extending the results of Scarf [44] and Azoury [4] using the

newsvendor distribution framework developed in Braden and Freimer [8]. Using an example with perishable inventory and exponentially-distributed demand with a gamma conjugate prior, they show that it is often optimal to "stalk" demand information by overstocking, and that a product can be a victim of its own success if its sales popularity prevents the retailer from maintaining enough stock to accurately assess true demand. Ding at el. [13] generalize these ideas using a Bayesian Markov decision process (BMDP) for a similar multi-period newsvendor problem setting, and obtain similar insights: stocking levels should be at least as high as those when lost sales are observed, and early period stocking levels should be higher to gather more information regarding true demand. Similar to the POMDP, the BMDP used by the researchers utilizes a probability-distribution-valued state variable, and thus has tractability issues for more complex problem settings. Finally, in another related work, Godrey and Powell [16] consider again the multi-period newsvendor setting with unobserved lost sales, and propose an approximate dynamic programming approach based on sampling to set nearly-optimal stock levels; the approach does not attempt to characterize the demand distribution, and focuses solely on generating near-optimal stock decisions by iteratively updating an approximation of the true expected profit objective function.

Here, we study a different, but related, problem class with non-perishable inventories. We assume that demands are non-stationary, but correlated across decision periods, such that the demand distribution in period $t+1$ depends on the demand level in period $t$, and thus we can model the demand process as a Markov chain. Furthermore, we assume that the process structure is known, *i.e.*, we know the transition matrix for the Markov chain. Since, however, we assume that lost sales are not observed, what is not known is the actual demand $d(t)$ in a period when a stockout occurs. Additionally, we assume we are also able to observe some additional market data, summarized by $z(t)$, that may help to identify $d(t)$. Since the relationship between $z(t)$ and $d(t)$ is only known probabilistically, we develop a POMDP model. Since the direct model is intractable for reasonable instance sizes, we consider heuristic solution algorithms similar to the limited look ahead policies proposed by Treharne and Sox [52] for a related problem with perfect (non-censored) observations of

demand and backordering. In their work, Treharne and Sox [52] assume that the demand in each period $t$ is generated by some demand distribution $d_t$ from a finite family, and that the distributions change from one period to the next according to a Markov chain with known transition probabilities. However, the state of the process $\{d_t\}$ is not observed directly at any time; instead, only the demand outcomes $\{w_t\}$ from distributions $\{d_t\}$ are observed. Unlike our work, the authors focus primarily on methods for determining near-optimal stocking policies for such environments, and do not address the value of improving the observability of the demand distribution in a period.

### *3.3 Inventory Control with Partially Observed Markovian Demand*

Consider a single product supply chain system where a single decision maker selects a replenishment quantity at each of a discrete, predefined and finite set of decision epochs in order to maximize expected total discounted profit over the problem horizon. Just prior to each decision epoch, the decision maker observes the current inventory level and the demand that has occurred since the last decision epoch. We assume that the inventory level is completely observed, but that the demand observation is potentially noise-corrupted, *i.e.*, obtained only via an observation of prior period sales and possibly augmented by noisy market signal data. Selection of the replenishment quantity at the current epoch is based on all past and present inventory and demand sensing observations and all past ordering decisions. We assume that the quantity ordered is received for use immediately and that demand is described by a control-independent (exogenous) Markov chain.

More precisely, let $x(t)$ be the completely observed inventory level at time (or decision epoch) $t$, just prior to the selection of the replenishment decision $a(t)$. Let $d(t)$ be the demand realized between time $t-1$ and time $t$. We assume replenishment decisions are made at each $t \in \{0, 1, \cdots, T-1\}$, where $T < \infty$; thus, the planning horizon is finite. Demand in excess of on-hand inventory is lost, and hence

$$x(t+1) = \max\{0, x(t) + a(t) - d(t+1)\}. \tag{2}$$

Note that these assumptions imply that sales are also observed, since the sales between

$t$ and $t+1$ are simply $x(t)+a(t)-x(t+1)$.

Let $z(t)$ be the noise-corrupted observation of demand (independent of sales) that is available just prior to the selection of $a(t)$, and assume probabilities of the form $P(z,j|i) = P(z(t+1)=z, d(t+1)=j|d(t)=i)$ are given. Note that we use $z(t)$ to summarize information that may be obtained from demand sensing programs. We note that $P(z,j|i) = P(z|j,i)P(j|i)$ where $P(j|i) = \sum_z P(z,j|i) = P(d(t+1)=j|d(t)=i)$ and

$$P(z|j,i) = \frac{P(z,j|i)}{P(j|i)} = P(z(t+1)=z|d(t+1)=j, d(t)=i),$$

assuming $P(j|i) \neq 0$. The probabilities $P(j|i)$ and $P(z|j,i)$ are appropriately referred to as *transition* and *observation* probabilities, respectively. We assume that $D$ is the maximum demand per period, and hence $d(t) \in \{0,1,\cdots,D\}$. Similarly, we assume $z(t) \in \{0,1,\cdots,D\}$.

We will have particular interest in two extreme cases of the observation probabilities. We note that $z(t) = d(t)$ $w.p.1$ for all $t$ is equivalent to $P(z|j,i) = 1$ if and only if $z = j$ for all $i$. In this case, we say that demand is completely (or perfectly) observed by the observation process $\{z(t), t = 1,2,\cdots\}$. If $P(z|j,i)$ is independent of $i$ and $j$, then the observation process provides no information about demand, and hence we say that demand is completely unobserved by the observation process. We remark that when the observation process provides no information about the demand process, information about the demand process can be inferred only from the inventory process, or equivalently, sales data, which will be described in section 3.4.

Selection of $a(t)$ is made with knowledge of the information set at time $t$, $\mathcal{I}(t)$, where $\mathcal{I}(t) = \Big\{z(t),\cdots,z(1),x(t),\cdots,x(0),a(t-1),\cdots,a(0),\xi(0)\Big\}$, $\xi(0) = \Big\{\xi_i(0)\Big\}$, and $\xi_i(0) = P(d(0)=i)$. Thus, $\xi(0) \in \Xi = \Big\{\xi \geq 0 : \sum_{i=0}^{D} \xi_i = 1\Big\}$. Hence, the amount of replenishment ordered at epoch $t$, $a(t)$, is allowed to depend on all past and present (possibly noise-corrupted) observations of demand, all past and present inventory levels, all former replenishment orders, and *a priori* demand information.

Let $p$, $\bar{p}$, $c$, and $h$ be the per unit selling price, salvage value, order cost, and per period inventory holding cost, respectively. We assume that holding cost from $t$ to $t+1$ is

determined on the basis of $x(t)$.

A policy $\pi$ is a rule that determines an action on the basis of the information currently available. Thus, $a(t) = \pi(t, \mathcal{I}(t))$ for all $t \in \{0, 1, \cdots, T-1\}$.

The *Inventory Replenishment Demand Sensing Problem* (*IRDSP*) is to find a policy that maximizes the following criterion with respect to all policies:

$$\mathbf{E}^{\pi}_{\xi(0)} = \left\{ \sum_{t=0}^{T-1} \beta^t r\left[s(t), a(t)\right] + \beta^T \bar{r}\left[s(T)\right] \right\}, \tag{3}$$

where $\mathbf{E}^{\pi}_{\xi(0)}$ is the expectation operator conditioned on $\xi(0)$ and use of policy $\pi$, $\beta$ is the discount factor, and where $s(t) = (x(t), d(t))$, $r\left[s(t), a(t)\right] = -hx(t) - ca(t) + p\mathbf{E}\Big\{ \min\{d(t+1), x(t) + a(t)\} \Big\}$ and $\bar{r}\left[s(T)\right] = \bar{p}x(T)$. Note that $\min\{d(t+1), x(t) + a(t)\}$ represents sales between $t$ and $t+1$.

It is easy to see that an optimal policy for the *IRDSP* will always select values of $a(t)$ such that $a(t) \leq D - x(t)$ given that there are no fixed ordering costs and no replenishment lead time in this problem setting.

### *3.4 Preliminary Results*

The following observations result from equation (2):

1. If $x(t+1) > 0$, then $d(t+1) = x(t) + a(t) - x(t+1)$, and hence $d(t+1)$ is completely observed.

2. If $x(t+1) = 0$, then all we can infer about $d(t+1)$ from $x(t+1)$, $x(t)$, and $a(t)$ is that $d(t+1) \geq x(t) + a(t)$.

It follows from Smallwood and Sondik [45] that $(x(t), \xi(t))$ represents a sufficient statistic for the *IRDSP*, where $\xi(t) = \{\xi_i(t)\} \in \Xi$ and $\xi_i(t) = P(d(t) = i | \mathcal{I}(t))$. This fact, coupled with the above two observations, imply that there are two general states of interest:

1. $(x, e_i)$, when $x > 0$, where the $j^{th}$ element of the vector $e_i$ is 1 if $i = j$ and 0 otherwise.

2. $(0, \xi)$ for any probability mass vector $\xi$ on the demand state.

Let

$$\tilde{\sigma}(z,(x,\xi),a) = \sum_{j \geq x+a} \sum_{i} \xi_i P(z,j|i)$$

$$\tilde{\lambda}_j(z,(x,\xi),a) = \begin{cases} 0 & j < x+a \\ \frac{\sum_i \xi_i P(z,j|i)}{\tilde{\sigma}(z,(x,\xi),a)} & j \geq x+a \end{cases},$$

where $\tilde{\sigma}(z,(x,\xi),a) \neq 0$, and $\tilde{\lambda}(z,(x,\xi),a) = \left\{\tilde{\lambda}_j(z,(x,\xi),a)\right\}$. Note that $\tilde{\sigma}(z,(x,\xi),a) = P\big(z(t+1) = z, x(t+1) = 0 | x(t) = x, \xi(t) = \xi, a(t) = a\big)$ and that $\tilde{\lambda}_j(z,(x,\xi),a) = P\big(d(t+1) = j | z(t+1) = z, x(t+1) = 0, x(t) = x, \xi(t) = \xi, a(t) = a\big)$. Thus, assuming $(x(t),\xi(t)) = (x,\xi)$:

1. if $x(t+1) > 0$, then $\xi(t+1) = e_i$, where $d(t+1) = x(t) + a(t) - x(t+1) = i$.

2. if $x(t+1) = 0$, then $\xi(t+1) = \tilde{\lambda}(z,(x,\xi),a)$ with probability $\tilde{\sigma}(z,(x,\xi),a)$, where $a(t) = a$ and $z(t+1) = z$.

Based on results in Smallwood and Sondik [45], we now develop optimality equations for the $x > 0$ and $x = 0$ cases. In both cases, $v_T(x,\xi) = \bar{p}x$.
If $x > 0$, then:

$$\begin{aligned} v_t(x,e_i) \;=\; \max_{a \geq 0} \Bigg\{ & -hx - ca + p \sum_j \min(j, x+a) P(j|i) \\ & + \beta \sum_{j < x+a} P(j|i) v_{t+1}(x+a-j, e_j) \\ & + \beta \sum_z \tilde{\sigma}(z,(x,e_i),a) v_{t+1}\left(0, \tilde{\lambda}(z,(x,e_i),a)\right) \Bigg\}. \end{aligned}$$

If $x = 0$, then:

$$\begin{aligned} v_t(0,\xi) \;=\; \max_{a \geq 0} \Bigg\{ & -ca + p \sum_i \xi_i \sum_j \min(j, a) P(j|i) \\ & + \beta \sum_{j < a} \sum_i \xi_i P(j|i) v_{t+1}(a-j, e_j) \\ & + \beta \sum_z \tilde{\sigma}(z,(0,\xi),a) v_{t+1}\left(0, \tilde{\lambda}(z,(0,\xi),a)\right) \Bigg\}. \end{aligned}$$

We observe that $\tilde{\sigma}$ and $\tilde{\lambda}$ depend on $x$ and $a$ only through $x+a$. For $y = x+a$, define:

$$\begin{aligned}
\sigma(z,\xi,y) &= \tilde{\sigma}(z,(x,\xi),a) \\
\lambda(x,\xi,y) &= \tilde{\lambda}(z,(x,\xi),a) \\
L(\xi,y) &= p\sum_i \xi_i \sum_j \min(j,y)P(j|i).
\end{aligned}$$

Note, $L(\xi,y) = \sum_i \xi_i L(e_i,y)$. Also let

$$\begin{aligned}
h(\xi,y,v) &= -cy + L(\xi,y) + \beta\sum_{j<y}\left[\sum_i \xi_i P(j|i)\right] v(y-j,e_j) \\
&\quad +\beta\sum_z \sigma(z,\xi,y)v\left(0,\lambda(z,\xi,y)\right)
\end{aligned}$$

and

$$[Hv](x,\xi) = (c-h)x + \max_{y\geq x} h(\xi,y,v).$$

Then, the optimality equation is $v_t = Hv_{t+1}$, where $v_T(x,\xi) = \bar{p}x$.

Results in Smallwood and Sondik [45] imply that:

1. $v_t(x,\xi)$ is the optimal expected reward to be accrued from $t$ until $T$, given $x(t) = x$ and $\xi(t) = \xi$.

2. An action that causes the maximum in the optimality equation to be attained is an optimal action for the concomitant state.

### *3.5 Numerical Algorithms*

We now present two general approaches for determining optimal policies and three heuristic algorithms for determining a sub-optimal policy for the *IRDSP* and the expected value accrued over the planning horizon. The approaches differ on the basis of the sufficient statistic used.

### 3.5.1 Approach 1

The first optimal algorithm uses $(x(t), \xi(t))$ as a sufficient statistic for $\mathcal{I}(t)$ and takes advantage of the fact that for each $t$, there is a finite set of vectors, $\Gamma_t$, such that $v_t(0,\xi) = \max\{\xi\gamma : \gamma \in \Gamma_t\}$; that is, $v_t(0,\xi)$ is piecewise linear and convex in $\xi$ for finite $T$. Thus, although the set of all probability mass vectors $\xi$ is uncountably infinite, $v_t(0,\cdot)$ has a finite representation (see Smallwood and Sondik [45]) .

Following arguments in Smallwood and Sondik [45], $\Gamma_t$ can be constructed from $\Gamma_{t+1}$ as follows. Note $v_T(0,\xi) = 0$ for all $\xi$; Thus, $\Gamma_T = \{0\}$. Then,

$$\begin{aligned} v_t(0,\xi) &= \max_{y\geq 0}\left\{-cy + L(\xi,y) + \beta\sum_{j<y}\left[\sum_i \xi_i P(j|i)\right] v_{t+1}(y-j,e_j)\right. \\ &\qquad \left. + \beta\sum_z \sigma(z,\xi,y)\max\{\lambda(z,\xi,y)\gamma : \gamma\in\Gamma_{t+1}\}\right\} \\ &= \max_{y\geq 0}\max_{\gamma^0}\cdots\max_{\gamma^Z}\left\{-cy + L(\xi,y) + \beta\sum_{j<y}\left[\sum_i \xi_i P(j|i)\right] v_{t+1}(y-j,e_j)\right. \\ &\qquad \left. + \beta\sum_z \sigma(z,\xi,y)\lambda(z,\xi,y)\gamma^z\right\}. \end{aligned}$$

It follows that $\sigma(z,\xi,y)\lambda(z,\xi,y)\gamma^z = \sum_{j\geq y}\left[\sum_i \xi_i P(z,j|i)\right]\gamma_j^z$, and hence

$$\begin{aligned} v_t(0,\xi) &= \max_{y\geq 0}\max_{\gamma^0}\cdots\max_{\gamma^Z}\left\{\sum_i \xi_i\left[-cy + L(e_i,y) + \beta\sum_{j<y} P(j|i) v_{t+1}(y-j,e_j)\right.\right. \\ &\qquad \left.\left. + \beta\sum_z\sum_{j\geq y} P(z,j|i)\gamma_j^z\right]\right\}. \end{aligned}$$

Thus, $\Gamma_t$ is composed of vectors $\gamma' = \{\gamma_i'\}$ of the form,

$$\gamma_i' = -cy + L(e_i,y) + \beta\sum_{j<y} P(j|i) v_{t+1}(y-j,e_j) + \beta\sum_z\sum_{j\geq y} P(z,j|i)\gamma_j^z.$$

We observe that if all of the vectors $\gamma'$ are contained in $\Gamma_t$, then $|\Gamma_t| = (D+1)\times|\Gamma_{t+1}|^{(D+1)}$; hence the finite representation of $v_t(0,\xi)$ expands geometrically as $T$ increases if no attempt

is made to remove unnecessary members of $\Gamma$. Let $Purge(\Gamma) \subseteq \Gamma$ be the subset of $\Gamma$ having the smallest cardinality such that $\max\{\xi\gamma : \gamma \in Purge(\Gamma)\} = \max\{\xi\gamma : \gamma \in \Gamma\}$ for all $\xi$. See Lin at el. [28] for results regarding the existence of $Purge(\Gamma)$ and computationally efficient ways to compute $Purge(\Gamma)$, given $\Gamma$.

This discussion suggests that a finite representation of $v_t$ is $(\hat{v}_t, \Gamma_t)$, where $v_t(x, e_i) = \hat{v}_t(x, i)$ for all $i$ and $x > 0$ and $v_t(0, \xi) = \max\{\xi\gamma : \gamma \in \Gamma_t\}$. Define the operators $H_1$ and $H_2$ as follows:

$$H_1(\hat{v}, \Gamma)(x, i) = [Hv](x, e_i),$$

for all $x > 0$ and all $i$, and

$$H_2(\hat{v}, \Gamma)(\xi) = [Hv](0, \xi),$$

for all $\xi$, where $v(x, e_i) = \hat{v}(x, i)$ for all $x > 0$ and all $i$ and $v(0, \xi) = \max\{\xi\gamma : \gamma \in \Gamma\}$ for all $\xi$. Then, $\hat{v}_t = H_1(\hat{v}_{t+1}, \Gamma_{t+1})$ and $\Gamma_t = H_2(\hat{v}_{t+1}, \Gamma_{t+1})$.

**Sub-Optimal Design**

Although the *Purge* operator can be useful, $|Purge(\Gamma_t)|$ may still grow prohibitively large as $T$ gets large. We now consider a sub-optimal design that guarantees the cardinality of $\Gamma_t$ will never exceed a computable upper bound.

Recalling that $D$ is the maximum demand, select $a(t) = D$, independent of $\xi$, at time $t$ if $x(t) = x(t-1) = \cdots x(t-K) = 0$ and $x(t-K-1) > 0$ for a fixed integer $K \geq 0$. Otherwise, select $a(t) \leq D$ that obtains the maximum value in the optimality equation. In the former case, once $a(t)$ is selected, the inventory level at the next decision epoch is guaranteed to be either greater than zero or a special case of zero inventory that allows complete demand observability. That is, note from equation 2 that if $a(t) = D$, $x(t) = 0$, and $x(t+1) = 0$, then $d(t+1) = D$.

We define:

- $v_t^K(x, e_i)$ as the expected reward to be accrued from $t$ until $T$ under the sub-optimal design policy with parameter $K$ given $\xi(t) = e_i$ and $x(t) = x$ where $x > 0$.

- $v_t^k(0,\xi) \quad \forall \quad k = 1, \cdots, K$ as the expected reward to be accrued from $t$ until $T$ under the sub-optimal design with parameter $K$ given $\xi(t) = \xi$, $x(t) = x(t-1) = \cdots = x(t-K-k) = 0$ and $x(t-K-k-1) > 0$.

- $v_t^0(0,\xi)$ as the expected reward to be accrued from $t$ until $T$ under the sub-optimal design with parameter $K$ given $\xi(t) = \xi$, $x(t) = x(t-1) = \cdots = x(t-K) = 0$.

- $\Gamma_t^k \quad \forall \quad k = 0, \cdots, K$ as the set of gamma vectors such that $v_t^k(0,\xi) = \max_{\gamma \in \Gamma_t^k}\{\xi\gamma\}$.

Assume $v_T^K(x,\xi) = \bar{p}x$ and hence $\Gamma_T^K = \{\gamma_T^0\}$, where $\gamma_T^0 = 0$. Furthermore, let

$$\gamma_{it}^0 = -cy^* + L(e_i, y^*) + \beta \sum_{j<y^*} P(j|i) v_{t+1}^K(y^* - j, e_j) + \beta \sum_{j \geq y^*} P(j|i) \gamma_{j\,t+1}^0. \tag{4}$$

Let $v_t^K = v_t,\ t = T-K, \cdots, T$.

**Algorithm 1**

For $t < T-K$, assume the array $(v_{t+k}^K,\ k = 1, \cdots K,\ \Gamma_{t+1}^K, \gamma_{t+K}^0)$ is given, where $v_{t+k}^K = \{v_{t+k}^K(x, e_i) : x > 0\},\ k = 1, \cdots, K$. We remark that this array fully determines $v_{t+1}^K(x,\xi)$ for all $x$ and $\xi$. We determine $(v_{t+k}^K,\ k = 0, \cdots K-1,\ \Gamma_t^K, \gamma_{t+K-1}^0)$ as follows:

(i) $\gamma_{t+K-1}^0$ is determined from $v_{t+K}^K$ and $\gamma_{t+K}^0$.

(ii) $v_t^K = H_1(v_{t+1}^K, \Gamma_{t+1}^K)$.

(iii) $\Gamma_{t+k}^{K-k} = H_2(v_{t+k+1}^K, \Gamma_{t+k+1}^{K-k-1})$ for $k = 0, \cdots, K-1$.

We note that the cardinality of the array $(v_{t+k}^K,\ k = 0, \cdots K-1,\ \Gamma_t^K, \gamma_{t+K-1}^0)$ is $K \times D \times (D+1) + |\Gamma_t^K| + 1$, where $|\Gamma_t^0| = 1$ and $|\Gamma_t^k| \leq (D+1) \times |\Gamma_{t+1}^{k-1}|^{(D+1)}$. Further, we note that transition from $(v_{t+1}^K, \cdots, \gamma_{t+K})$ to $(v_t^K, \cdots, \gamma_{t+K-1})$ requires application of the $H_2$ operator $K$ times.

We explain the use of the $\gamma_t^0$ vector as follows. For simplicity, let $K = 0$; that is, assume we order $y^*$ items whenever the inventory goes to zero, irrespective of $\xi$ (in reality not a particularly clever sub-optimal design). Let $v_t^0$ be the resulting expected value to be accrued from $t$ until $T$. Then,

$$
\begin{aligned}
v_t^0(0,\xi) &= -cy^* + L(\xi,y^*) + \beta \sum_{j<y^*} \left[ \sum_i \xi_i P(j|i) \right] v_{t+1}^0(y^*-j, e_j) \\
&\quad + \beta \sum_z \sigma(z,\xi,y^*) v_{t+1}^0(0, \lambda(z,\xi,y^*)).
\end{aligned}
$$

We recall that $v_T^0(0,\xi) = 0$; hence, $\Gamma_T^0 = \{0\}$. Assume $\Gamma_{t+1}^0$ is also a singleton; *i.e.*, $\Gamma_{t+1}^0 = \{\gamma_{t+1}^0\}$. Then,

$$
\begin{aligned}
\sum_z \sigma(z,\xi,y^*) v_{t+1}^0(0,\lambda(z,\xi,y^*)) &= \sum_z \sigma(z,\xi,y^*)\lambda(z,\xi,y^*)\gamma_{j\,t+1}^0 \\
&= \sum_z \sum_{j\geq y^*} \left[ \sum_i \xi_i P(z,j|i) \right] \gamma_{j\,t+1}^0 \\
&= \sum_{j\geq y^*} \left[ \sum_i \xi_i P(j|i) \right] \gamma_{j\,t+1}^0,
\end{aligned}
$$

where the last equality is due to the fact that $\sum_z P(z,j|i) = P(j|i)$ and that $\gamma_{t+1}^0$ is independent of $z$. Thus, if $\Gamma_{t+1}^0$ is a singleton and the action taken is $\xi$-invariant, then $\Gamma_t^0$ is also an (easily computed) singleton.

It seems reasonable that $v_t^{K+1}$ would be at least as good an approximation as $v_t^K$, which we now show.

**Proposition 2** *For all t,* $v_t^K \leq v_t^{K+1} \leq v_t$.

**Proof:** Clearly, for any $K$, $v_t^K \leq v_t$ for all $t$. Hence, it is sufficient to show that $v_t^K \leq v_t^{K+1}$. By definition $v_{T-k}^K = v_{T-k}^{K+1} = v_{T-k}$ for $k = 0, \cdots, K$, and $v_{t-k}^{K+1} = v_{T-k}$ for $k = K+1$. Thus, $v_t^K \leq v_t^{K+1}$ for $t = T-K-1$. For any $t < T-K-1$, assume $v_{t+k}^K \leq v_{t+k}^{K+1}$, for $k = 1, \cdots, T-t$. The monotonicity of the operators $H_1$ and $H_2$ guarantee that $H_1(v_{t+1}^K, \Gamma_{t+1}^K) \leq H_1(v_{t+1}^{K+1}, \Gamma_{t+1}^{K+1})$ and $v_t^K(0,\xi) \leq v_t^{K+1}(0,\xi)$ for all $\xi$, if $v_{t+1}^{K-1}(0,\xi) \leq v_{t+1}^K(0,\xi)$ for all $\xi$. It is then straightforward to show that $v_{t+K}^0(0,\xi) \leq v_{t+K}^1(0,\xi)$. The monotonicity of $H_2$ then implies $v_{t+1}^{K-1} \leq v_{t+1}^K$, and the result follows by induction.

■

We now present an alternative approach for determining $v_t^K$.

**Algorithm 2**

For $t < T-K$ assume the array $(v_{t+1}^K, \Gamma_{t+1}^k, k = 0, \cdots, K)$ is given, where $v_{t+1}^K = \{v_{t+1}^K(x, e_i) : x > 0\}$. We remark that this array fully determines $v_{t+1}^K(x, \xi)$ for all $x$ and $\xi$. We determine $(v_t^K, \Gamma_t^k, k = 0, \cdots, K)$ as follows:

(i) $\gamma_t^0$ is determined from $v_{t+1}^K$ and $\gamma_{t+1}^0$.

(ii) $v_t^K = H_1(v_{t+1}^K, \Gamma_{t+1}^K)$.

(iii) $\Gamma_t^k = H_2(v_{t+1}^K, \Gamma_{t+1}^{k-1})$ for $k = 1, \cdots, K$.

We note that the cardinality of the array $(v_t^K, \Gamma_t^k, k = 0, \cdots, K)$ is $D \times (D+1) + \sum_{k=0}^K |\Gamma_t^k|$, where $|\Gamma_t^0| = 1$ and $|\Gamma_t^k| \leq (D+1) \times |\Gamma_{t+1}^{k-1}|^{(D+1)}$. Further, we note that transition from $(v_{t+1}^K, \Gamma_{t+1}^k, k = 0, \cdots, K)$ to $(v_t^K, \Gamma_t^k, k = 0, \cdots, K)$ requires application of the $H_2$ operator $K$ times.

We remark that on the basis of operations count, Algorithm 1 would be preferred to, Algorithm 2. However, as we will now show, Algorithm 2 suggests an algorithm, Algorithm 3 presented below, that is based in a non-standard sufficient statistic offering a significantly simpler approach for software development.

### 3.5.2 Approach 2

The first approach for constructing an optimal policy for the *IRDSP* was based on the fact that for finite $T$, $v_t(0, \xi)$ has a finite representation, $\Gamma_t$, although $\xi$ is a member of an uncountably infinite set. The second approach for constructing an optimal policy is based on the fact that $|\mathcal{I}(t)|$ is finite for finite $t$. We also make use of the fact that there exists a set $\mathcal{I}' \subseteq \mathcal{I}(t)$ that can also serve as a sufficient statistic for the *IRDSP*, where $\mathcal{I}'(t) = \{z(t), \cdots, z(t-\tau+1), a(t-1), \cdots, a(t-\tau), x(t-\tau), d(t-\tau)\}$, and where $\tau$ is such

that $x(t) = x(t-1) = \cdots = x(t-\tau+1) = 0$ and $x(t-\tau) > 0$. Proof of the following result, which justifies the claim that $\mathcal{I}'$ is a sufficient statistic, is due to the fact that $x(t) > 0$ implies $d(t)$ is completely observed.

**Proposition 3** *For all t,* $P(d(t) = i|\mathcal{I}'(t)) = P(d(t) = i|\mathcal{I}(t))$.

Let $\mathcal{I}_0 = \big\{(x, e_i) : x > 0, i \in \{0, 1, \cdots, D\}\big\}$, $\mathcal{I}_1 = \{\lambda(z, \xi, y) : (x, \xi) \in \mathcal{I}_0, y \in \{x, x+1, \cdots, D\}, z \in \{0, 1, \cdots, D\}\}$ and for $k \geq 1$ let $\mathcal{I}_{k+1} = \big\{\lambda(z, \xi, y) : \xi \in \mathcal{I}_k,\ z, y \in \{0, \cdots, D\}\big\}$. We remark that $\mathcal{I}_k$ is equivalent to $\mathcal{I}'(t)$, given $\tau = k$. Note, that $|\mathcal{I}_0| = D(D+1)$ and $|\mathcal{I}_k| \leq \frac{D(D+1)^{2k+1}}{2}$ for $k \geq 1$. As a slight abuse of notation, let $H_1(v, \tilde{v}) = H_1(v, \Gamma)$ and $H_2(v, \tilde{v}) = H_2(v, \Gamma)$ if $\tilde{v}(0, \xi) = \max\{\xi\gamma : \gamma \in \Gamma\}$. We now present an algorithm for determining $v_t^K(x, \xi)$ for all $(x, \xi) \in \mathcal{I}_0$ and $v_t^{K-k+1}(0, \xi)$ for all $k = 1, \cdots, K+1$ and for all $(0, \xi)$ such that $\xi \in \mathcal{I}_k$.

**Algorithm 3**

For $t < T - K$, assume $v_{t+1}^K(x, \xi)$ is given, for all $(x, \xi) \in \mathcal{I}_0$ and $v_t^{K-k+1}(0, \xi)$ is given for all $k = 1, \cdots, K+1$ and for all $(0, \xi)$ such that $\xi \in \mathcal{I}_k$, where $v_{t+1}^0(0, \xi) = \xi\gamma_{t+1}^0$ for all $\xi \in \mathcal{I}_{K+1}$ and $\gamma_{t+1}^0$ is given. Assume $\overline{v}_{t+1} = \{v_{t+1}^K(x, \xi) : (x, \xi) \in \mathcal{I}_0\}$. We determine $v_t^K(x, \xi)$ for all $(x, \xi) \in \mathcal{I}_0$ and $v_t^{K-k+1}(0, \xi)$ for all $k = 1, \cdots, K+1$ and for all $(0, \xi)$ such that $\xi \in \mathcal{I}_k$, where $v_t^0(0, \xi) = \xi\gamma_t^0$ for $\xi \in \mathcal{I}_{K+1}$ as follows:

(i) $v_t^K(x, \xi) = H_1(\overline{v}_{t+1}, v_{t+1}^K(0, \cdot))(x, \xi)$ for all $(x, \xi) \in \mathcal{I}_0$ where
$v_{t+1}(0, \cdot) = \{v_{t+1}^K(0, \xi) : \xi \in \mathcal{I}_1\}$.

(ii) $v_t^{K-k+1}(0, \xi) = H_2(\overline{v}_{t+1}, v_{t+1}^{K-k}(0, \cdot))(0, \xi)$ for all $\xi \in \mathcal{I}_k$, where
$v_{t+1}^{K-k}(0, \cdot) = \{v_{t+1}^{K-k}(0, \xi) : \xi \in \mathcal{I}_{k+1}\}, k = 1, \cdots, K$.

(iii) $v_t^0(0, \xi) = \xi\gamma_t^0$, where
$\gamma_{it}^0 = -cy^* + L(e_i, y^*) + \beta \sum_{j<y^*} P(j|i) v_{t+1}^K(y-j, e_j) + \beta \sum_{j \geq y^*} P(j|i)\gamma_{j\,t+1}^0$.

We remark that Algorithm 2 and 3 are nearly identical, differing only as follows:

(i) The algorithms use different representations of the $v_t^k(0,\cdot)$ functions, where the representation in Algorithm 3 is significantly simpler for software implementation than is the representation in Algorithm 2.

(ii) Algorithm 3 holds for all $(x,\xi) \in \mathcal{I}_0$ and all $(0,\xi)$ such that $\xi \in \mathcal{I}_1 \bigcup \cdots \bigcup \mathcal{I}_K$, whereas Algorithm 2 holds for all $(x,\xi) \in \mathcal{I}_0$ and for all $(0,\xi)$ such that $\xi \in \Xi$.

### *3.6 A Method for Bounding the Value of Demand Observability*

We now present a procedure that uses the previously described model and solution approaches to bound the maximum value of improved demand observability for the *IRDSP*. To do so, we consider two extreme cases that we call *completely-observed* and *sales-only-observed*. In the completely-observed case, we assume that the observation process provides a perfect observation of demand in the prior period, even when $x(t) = 0$. Let $V_o^*(x,\xi)$ denote the value of maximum expected profit over some fixed planning horizon for this case, given $x(0) = x$ and $\xi(0) = \xi$. At the other extreme, the sales-only-observed case assumes that the observation process provides no additional information about demand. Therefore, the decision maker bases his or her decision only on the information obtained from sales data. Since this case corresponds to the situation in which the use of demand sensing techniques to improve observability provides no benefit, it should be useful for developing a lower bound. Let $V_s^*(x,\xi)$ denote the value of maximum expected profit in this case, again given $x(0) = x$ and $\xi(0) = \xi$. Results in White and Harrington [55] guarantee that $V_o^*(x,\xi) \geq V_s^*(x,\xi)$, for all $(x,\xi)$.

Clearly, the gap between these two values, $V_o^*(x,\xi) - V_s^*(x,\xi)$ corresponds to the *maximum* added expected benefit that can result through the application of techniques that aim at improving demand observability.

As described earlier, the complicating feature of the operator $H$ is due to the partial observability of the demand. Therefore, determining an optimal policy and the resultant maximum expected profit for the completely-observed case does not require the use of sub-optimal techniques for problem settings of reasonable size. On the other hand, it will usually be computationally prohibitive to determine an optimal policy for the sales-only-observed

case, so instead we turn to the suboptimal solution approaches developed in Section 3.5. Let $V_s^{LB(K)}(x,\xi)$ denote a lower bound for $V_s^*(x,\xi)$, obtained by applying the suboptimal design with parameter $K$. Thus,

$$V_o^*(x,\xi) - V_s^{LB(K)}(x,\xi) \tag{5}$$

corresponds to an upper bound on the maximum added value that can be obtained from improving demand observability. Of course, larger values of $K$ lead to tighter lower bounds $V_s^{LB(K)}(x,\xi)$ which in turn lead to tighter upper bounds on the maximum value due to improved demand observability.

### 3.6.1 Completely-Observed Case

To model the completely-observed case of the *IRDSP*, we simply assume that $P(z|j,i) = 1$ if and only if $z = j$. Thus, $z(t)$ is a perfect observation of $d(t)$ (independent of the value of $x(t)$). Hence, the only general state of interest is now $(x, e_i)$, $x \geq 0$. Let

$$h'(i,y,v) = -cy + L(e_i,y) + \sum_{j<y} P(j|i)v(y-j,j) + \sum_{j\geq y} P(j|i)v(0,j)$$

and

$$[H'v'](x,i) = (c-h)x + \max_{y\geq x} h'(i,y,v).$$

Then,

$$v'_t = H'v'_{t+1} \tag{6}$$

is the optimality equation for the completely-observed case where the boundary condition is $v'_T(x,i) = (\bar{p})x$.

Noting results in Cheng and Sethi [11] and elsewhere, it is easy to show that there exists an optimal policy in this case that is an order-up-to-policy. That is, there are integers $\{y_i^*(t)\}$ such that

$$a(t) = \begin{cases} y_i^*(t) - x & if \quad x \leq y_i^*(t) \\ 0 & otherwise \end{cases}$$

represents an optimal action at time $t$, where $x(t) = x$ and $d(t) = i$.

### 3.6.2 Sales-Only-Observed Case

To model the case where demand is only observable through sales data, we assume that $P(z|j,i)$ is independent of $i$ and $j$. Thus, $z(t)$ contains no information about the value of $d(t)$. The general optimality equation now becomes $v_t'' = H''v_{t+1}''$, $v_T''(x,\xi) = \bar{p}x$, where:

$$H''v'' = (c-h)x + \max_{y \geq x} h''(x,\xi,y,v),$$

$$h''(x,\xi,y,v) = -cy + L(\xi,y) + \sum_{j<y} \left[ \sum_i \xi_i P(j|i) \right] v(y-j, e_j) + \sigma''(\xi,y) v(0, \lambda''(\xi,y)),$$

$\sigma''(\xi,y) = \sum_{j \geq y} \sum_i \xi_i P(j|i)$ and for $\sigma''(\xi,y) \neq 0$,

$$\lambda_j''(\xi,y) = \begin{cases} 0 & j < y \\ \frac{\sum_i \xi_i P(j|i)}{\sigma''(\xi,y)} & j \geq y. \end{cases}$$

### 3.6.3 Computing a Bound

For the completely-observed case, we use recursive expression (6) to determine $V_o^*(x,\xi) = v_0'(x,\xi)$ for all potential initial states. It is important to note that these states include all $(x,\xi) \in \mathcal{I}_0$, as well as states $(0,e_i)$ for $i = 0,1,\cdots,D$. For the sales-only-observed case, we use suboptimal Algorithm 3 to determine $V_s^{LB(K)}(x,\xi) \leq v_0''(x,\xi)$ because it is easier to implement in software relative to Algorithms 1 and 2. In this case, the potential initial states again include all $(x,\xi) \in \mathcal{I}_0$, but we restrict the zero inventory states to those that might be visited by Algorithm 3; *i.e.*, , all states $(0,\xi)$ such that $\xi \in \mathcal{I}_1'' \bigcup \cdots \bigcup \mathcal{I}_{K+1}''$, where $\mathcal{I}_1'' = \{\lambda''(\xi,y) \ \ \forall\ (x,\xi) \in \mathcal{I}_0,\ y \in \{x, x+1, \cdots, D\}\}$ and $\mathcal{I}_{k+1}'' = \{\lambda(\xi,y) \ \ \forall\ \xi \in \mathcal{I}_k,\ y \in \{0,1,\cdots,D\}\}$ for $1 \leq k \leq K$.

Given an initial state $(x, \xi) \in \mathcal{I}_0$ that is shared by both cases, it is possible to calculate a bound on the expected added benefit of demand observability using $V_o^*(x,\xi) - V_s^{LB(K)}(x,\xi)$. For initial states $(0, \xi)$, a similar bound can be computed by blending the value function using the prior distribution. To do so, let $V_o^*(0,\xi) = \sum_i \xi_i v_0'(0, e_i)$. Then $V_o^*(0,\xi) - V_s^{LB(K)}(0,\xi)$ again represents a bound.

Since it may be useful to determine a measure for the potential value of observability that is independent of the initial state, we note that a reasonable approach may be to compare a weighted sum of the state-wise maximum percentage potential gains due to improved demand observability, where the weights given to each state correspond to its likelihood of visitation. Given scalar weights $w^\xi \geq 0$ for all $\xi \in \mathcal{I}_1'' \bigcup \cdots \bigcup \mathcal{I}_{K+1}''$ and $w^{x,\xi} \geq 0$ for all $(x,\xi) \in \mathcal{I}_0$ such that $\sum w^\xi + \sum w^{x,\xi} = 1$, such a weighted statistic is given by the following expression:

$$\begin{aligned} G^{UB(K)}(w) \quad = \quad & \sum_{\xi \in \mathcal{I}_1'' \bigcup \cdots \bigcup \mathcal{I}_{K+1}''} w^\xi \frac{\left(V_o^*(0,\xi) - V_s^{LB(K)}(0,\xi)\right)}{V_s^{LB(K)}(0,\xi)} \\ & + \sum_{(x,\xi) \in \mathcal{I}_0} w^{x,\xi} \frac{\left(V_o^*(x,\xi) - V_s^{LB(K)}(x,\xi)\right)}{V_s^{LB(K)}(x,\xi)}. \end{aligned} \tag{7}$$

### *3.7 Computational Analysis*

We now apply our approach for bounding the value of demand observability to a set of example finite horizon problem scenarios. The goal of the analysis presented below is to develop a better understanding of the impact of two important problem features on the value of observability: (1) the characteristics of the stochastic demand process, and (2) the relative contribution of holding cost to overall supply chain cost. Scenarios are generated by varying these two features, while holding remaining problem parameters constant.

Scenarios were generated for analysis using the following parameters, chosen to be representative of many typical real-world supply networks. The per unit ordering cost and selling price are set to be $c = 12$ and $p = 14$, and the end-of-horizon inventory salvage value $\bar{p}$ was set to zero. The per period holding cost $h$ is varied between scenarios, and is selected

from the set $\{0.2, 0.3, 0.5, 1, 2\}$. Note that in the scenarios with the smallest holding cost ($h = 0.2$), the per unit product contribution $p - c = 2$ erodes to zero after holding the unit for 10 periods and that in the scenarios with highest holding cost ($h = 2$), this contribution erodes after only holding the unit for one period in inventory. The discount factor chosen was $\beta = 0.95$.

To mitigate initial effects, we used a long maximum planning horizon of $T = 1000$ periods for each scenario. In practice, the value function estimates tend to converge much more rapidly. For the recursive algorithms used to solve the completely-observed and sales-only-observed special cases, we use the following stopping criterion: stop at iteration $t^*$, where $t^*$ is the minimum $t$ satisfying $1 \leq t \leq T$ and is such that the maximum state-wise absolute difference between the expected profit to be accrued from $T - t$ to $T$ and the expected profit to be accrued from $T - t + 1$ to $T$ is less than or equal to a predefined value $\epsilon$ (*i.e.*, $t^*$ is the minimum $t$ such that $\max_{x,\xi} |v^*_{T-t+1}(x, \xi) - v^*_{T-t}(x, \xi)| \leq \epsilon$). If such $t^*$ does not exist, stop at iteration $T$. In our computational study we set $\epsilon = 10^{-6}$. The maximum number of iterations required to solve the scenarios presented in this section was 326 ($< T$). Therefore, it is also true that the stationary policies (and their resultant state-wise expected profits) obtained in the final iteration are good approximations of the infinite horizon version of the problem.

Each scenario uses a Markovian demand process dependent on two parameters, $\zeta$ and $r$. Parameter $\zeta$ is the probability that the demand level in period $t + 1$ is the same as the demand in period $t$. The maximum amount by which the demand level may change from one period to the next is a function of parameter $r$; that is, $d(t + 1)$ takes values in the interval $[\max\{0, d(t) - r\}, \min\{D, d(t) + r\}]$, where $r$ is assumed to be less than or equal to $\frac{D}{2}$. For $1 \leq d(t) \leq D - 1$ we assume that $P(d(t + 1) > d(t)) = P(d(t + 1) < d(t)) = \frac{1-\zeta}{2}$. When $d(t) = 0$ the probability of an increment in demand in period $t+1$ is $1-\zeta$. Similarly when $d(t) = D$ the probability of a decrement in demand in period $t + 1$ is $1 - \zeta$.

If $d(t) - r \geq 0$ and $d(t) + r \leq D$ then demand in period $t+1$ has a symmetrical discrete triangular distribution around $d(t)$ according to expression:

$$P(d(t+1) = d(t) + j) = P(d(t+1) = d(t) - j) = \frac{(1-\zeta)}{r(r+1)}(r+1-j) \tag{8}$$

for $1 \leq j \leq r$.

If $d(t) = 0$ then demand in period $t+1$ has a discrete triangular distribution according to expression:

$$P(d(t+1) = j) = 2\frac{(1-\zeta)}{r(r+1)}(r+1-j)$$

for $1 \leq j \leq r$. Similarly, when $d(t) = D$ then

$$P(d(t+1) = D - j) = 2\frac{(1-\zeta)}{r(r+1)}(r+1-j)$$

for $1 \leq j \leq r$.

If $1 \leq d(t) \leq r-1$ then demand in period $t+1$ has a discrete triangular distribution around $d(t)$ where

$$P(d(t+1) = d(t) - j) = \frac{(1-\zeta)}{d(t)(d(t)+1)}(d(t)+1-j)$$

for $1 \leq j \leq d(t)$ and right tail distributed according to expression (8). On the other hand if $D-r+1 \leq d(t) \leq D$, then the demand in period $t+1$ has a discrete triangular distribution around $d(t)$ where

$$P(d(t+1) = d(t) + j) = \frac{(1-\zeta)}{(D-d(t))(D-d(t)+1)}(D-d(t)+1-j)$$

for $1 \leq j \leq d(t)$ and left tail distributed according to expression (8). Note that the probability that demand changes decreases as the magnitude of the change increases.

By varying parameters $\zeta$ and $r$, we can control the *volatility* of the demand process. We define $P(\zeta, r)$ to be at least as *volatile* as $P(\zeta', r')$ if and only if $\zeta \leq \zeta'$ and $r \geq r'$. As the value of $\zeta$ approaches 1 and the value of $r$ approaches 0, the process is less volatile. In our computational experiments, we consider various levels of demand volatility in order to observe its effect on the value of demand observability. To generate scenarios with different demand volatility for each inventory holding cost, we develop a separate scenario

with demand process $P(\zeta, r)$ for each combination of parameter values $r \in \{1, 2, 3, 4, 5\}$ and $\zeta \in \{0.6, 0.7, 0.8, 0.9, 0.925, 0.95, 0.975, 0.99\}$. Further, we set $D = 10$.

For each problem scenario, we calculate state-wise expected profits for the completely-observed case $V_o^*(x, e_i)$ for all $x \geq 0$ and $i \in \{0, \cdots, D\}$ using equation 6. Next, we use suboptimal Algorithm 2 with parameter $K = 2$ to determine $V_s^{LB(2)}(x, \xi)$ for all $(x, \xi) \in \mathcal{I}_0$ and $(0, \xi)$ such that $\xi \in \mathcal{I}_1'' \bigcup \cdots \bigcup \mathcal{I}_{K+1}''$. For all $(x, \xi) \in \mathcal{I}_0$ and $(0, \xi)$ such that $\xi \in \mathcal{I}_1'' \bigcup \cdots \bigcup \mathcal{I}_{K+1}''$ let:

- $a_t^{LB}(x, \xi)$ be the decision rule at time $t$ found using the sub-optimal design in the sales-only-observed case, given $x(t) = x$ and $\xi(t) = \xi$;
- $P^t_{(x,\xi),(\overline{x},\overline{\xi})}(a_t^{LB}(x, \xi)) = P(x(t+1) = \overline{x}, \xi(t+1) = \overline{\xi} | x(t) = x, \xi(t) = \xi, a(t) = a_t^{LB}(x, \xi))$ be the transition probability from state $(x, \xi)$ to state $(\overline{x}, \overline{\xi})$ at time $t$ given the sub-optimal decision rule;
- and $P(t) = \{P^t_{(x,\xi),(\overline{x},\overline{\xi})}(a_t^{LB})\}$ be the corresponding transition probability matrix for time $t$ given the sub-optimal decision rule.

As mentioned earlier, due to the length of the planning horizon and the utilized stopping criterion, a good approximation of the steady-state probabilities for all $(x, \xi) \in \mathcal{I}_0$ and $(0, \xi)$ such that $\xi \in \mathcal{I}_1'' \bigcup \cdots \bigcup \mathcal{I}_{K+1}''$ is obtained by calculating the steady-state probabilities of the Markov chain associated with the stochastic matrix $P(0)$. We then use these steady-state probabilities as the weights $w^{\xi}$ to compute the maximum percentage value statistic given by (7).

Before we present results on the maximum value of observability, it is first interesting to briefly discuss the characteristics of the state-dependent replenishment quantities $a_0$ determined by suboptimal Algorithm 2 for the sales-only-observed case. Of particular interest are the quantities determined for states in which a stockout has occurred, *i.e.*, $(x, \xi) = (0, \xi)$ where $\xi \in \mathcal{I}_1'' \bigcup \cdots \bigcup \mathcal{I}_{K+1}''$. For the scenarios considered in this study, there is evidence that the average order quantity tends to increase as the number of stockouts in a row increases. One approach to display this phenomenon graphically is to consider, for a

given scenario, the order quantities selected in states $(0, \xi) \in \mathcal{I}_1''$ versus the order quantities selected in comparable states $(0, \xi) \in \mathcal{I}_2''$. Since the two sets $\mathcal{I}_1''$ and $\mathcal{I}_2''$ will for the most part consist of different elements $\xi$, we define two states $(0, \xi_1)$ and $(0, \xi_2)$ to be comparable if $E[\xi_1] = E[\xi_2]$. Figure 8 provides such comparative plots of $a_0(0, \xi)$ versus $E[\xi]$, for the scenario with $h = 0.5$, $\zeta = 0.7$, and $r = 3$. In this figure, note that for any $\xi \in \mathcal{I}_1''$ and for any $\xi' \in \mathcal{I}_2''$ such that $a_0(0, \xi) = a_0(0, \xi')$ it can be observed that $E(\xi') \leq E(\xi)$. Such ordering patterns are quite intuitive: when an order quantity of a certain level implemented after occurrence of a stockout results in yet another stockout, the quantity should be increased (given positively correlated demand) to increase the likelihood that demand may be served in the following period. Another way to interpret this behavior is that increasing the order quantity for a given expected demand level each time a stockout occurs is an attempt to "stalk" information about true demand and improve profitability; note that this argument does not depend on the demand correlation structure.

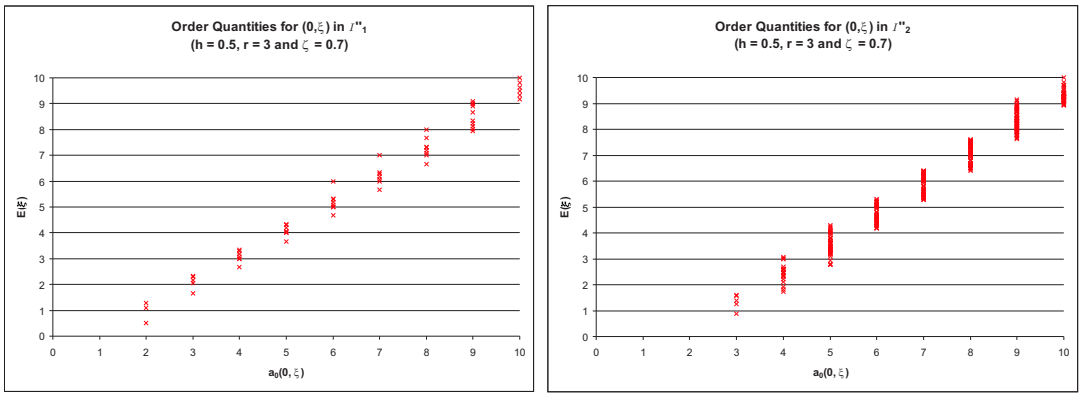



**Figure 8:** Comparison of Order Quantities for Stockout States in the Sales-Only-Observed Case: $\mathcal{I}_1''$ versus $\mathcal{I}_2''$ for $h = 0.5$, $\zeta = 0.7$, and $r = 3$

We now summarize the results of the experiments. First, it is important to mention that the maximum percentage value statistics calculated using (7) varied widely across the scenarios, from a minimum of approximately 2% to a maximum of nearly 35%. To develop an understanding of how the parameters of a scenario affected this measure, we begin by discussing the impact of demand volatility. In general, scenarios with low demand volatility benefit more from demand observability than scenarios with higher demand volatility. This

result is quite intuitive, since it is natural to believe that demand sensing may be an effective approach for improving profitability when demand is fairly stable and therefore more predictable. Of course, there is a limit to this argument since when demand is *completely* predictable, there should be very little additional value to increasing its observability.

Figures 9, 10, and 11 present the maximum value of demand observability statistics for three different holding cost levels, across a wide range of demand parameters. Figure 9 presents results for scenarios with a low holding cost value $h = 0.2$, Figure 10 presents results for a medium holding cost value $h = 0.5$, and Figure 11 presents results for a high holding cost value $h = 1.0$. We can observe from these figures that in general, for a constant value of $r$ the percentage increase in profitability increases as the value of $\zeta$ increases. It is also interesting to observe that for a fixed value of $h$ and high values of $\zeta$ (*i.e.*, $\zeta \geq 0.9$), the value of demand observability increases as $r$ decreases. Thus, in these cases, lower demand volatility leads to greater value of observability.

Figures 9, 10, and 11 also show that decreasing volatility does not always lead to an increase in the relative value of observability. For example, in the case with the lowest holding cost $h = 0.2$, the value of observability tends to grow slightly with $r$ (and thus increases with increasing volatility) for the lowest values of $\zeta$. Figure 12 graphically depicts the results for scenarios with medium holding cost $h = 0.5$, and it is easy to see that increasing volatility due to increasing values of parameter $r$ has different effects on the value of observability for different values of $\zeta$. There is also evidence in other scenarios that decreasing volatility due to increasing $\zeta$ does not always lead to increasing value of observability. Figure 13 depicts the weighted percentage gains as a function of $\zeta$ for a fixed low holding cost $h = 0.2$ and fixed $r = 2$. While the potential gain increases with $\zeta$ at lower values of $\zeta$, it does begin to decrease again for the highest value of $\zeta$.

Next, we investigate the impact of the relative contribution of holding cost to total system cost on the maximum value of improving demand observability. The computational results indicate that the value of the per unit per period holding cost $h$ has a significant effect on the value of observability. In general, as $h$ increases the value increases when all other parameters are held constant. Figures 14 and 15 graphically depict this observation for all

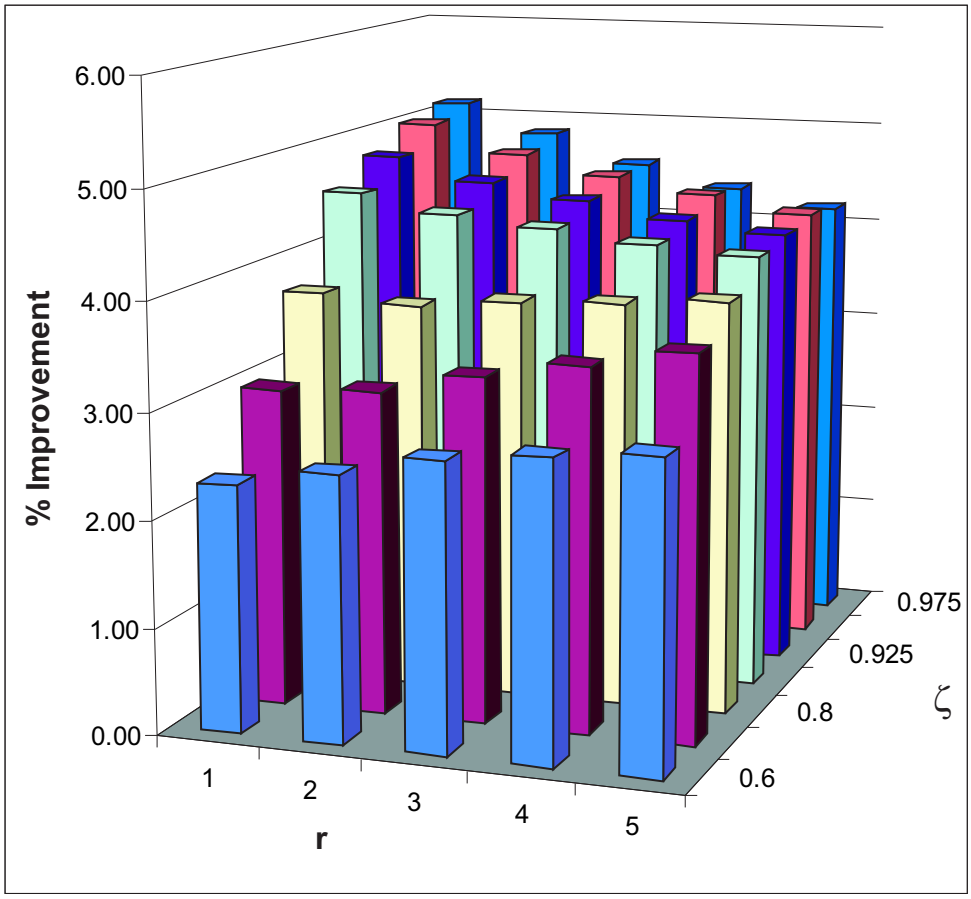



**Figure 9:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for Low Holding Cost Scenarios ($h = 0.2$)

scenarios with $r = 3$. That increased holding cost rates lead to an increase in the value of demand observability is again consistent with intuition. Increased holding cost rates tend to drive a system to carry less inventory, and systems with less inventory are more likely to incur stockouts that screen observations of true demand when only sales are observed. Averaging the potential percentage gains over all scenarios with low inventory holding cost rates $h \in \{0.2, 0.3, 0.5\}$, we find an average value of observability of about 5%. When averaging over scenarios with high inventory cost rates $h \in \{1.0, 2.0\}$, the average value is about 13%. It is important to note, however, that our state-wise bounds on the maximum value of observability are likely to be looser for high values of $h$ than for low values. The suboptimal algorithm used to determine $V_s^{LB(2)}$ will clearly lead to worse results when the inventory holding cost rate is higher, since the policy of ordering the maximum demand $D$

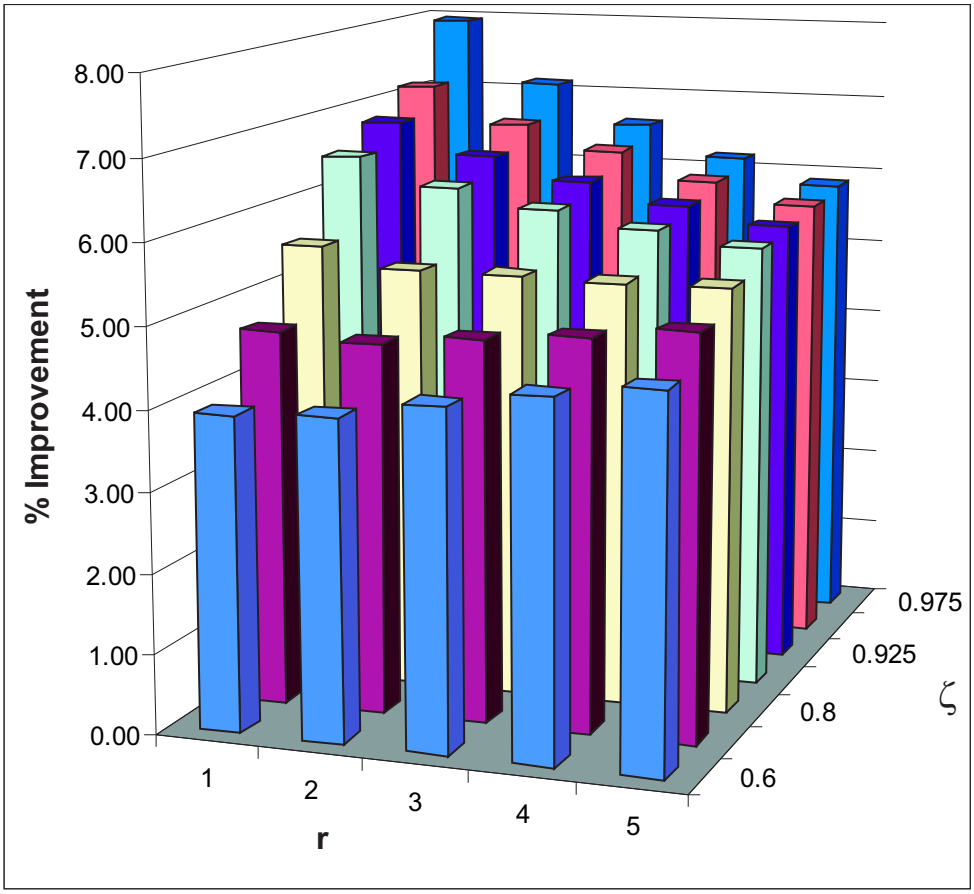



**Figure 10:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for Medium Holding Cost Scenarios ($h = 0.5$)

once we have observed three consecutive stockouts may lead to high holding costs.

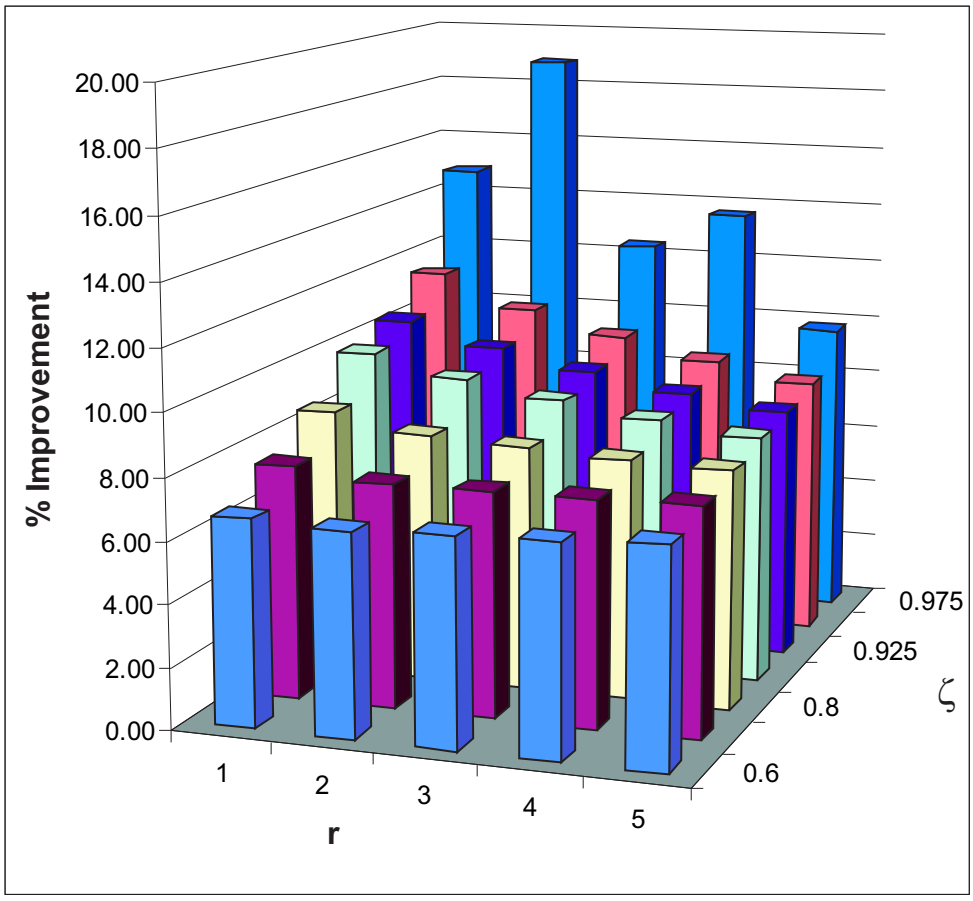



**Figure 11:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for High Holding Cost Scenarios ($h = 1.0$)

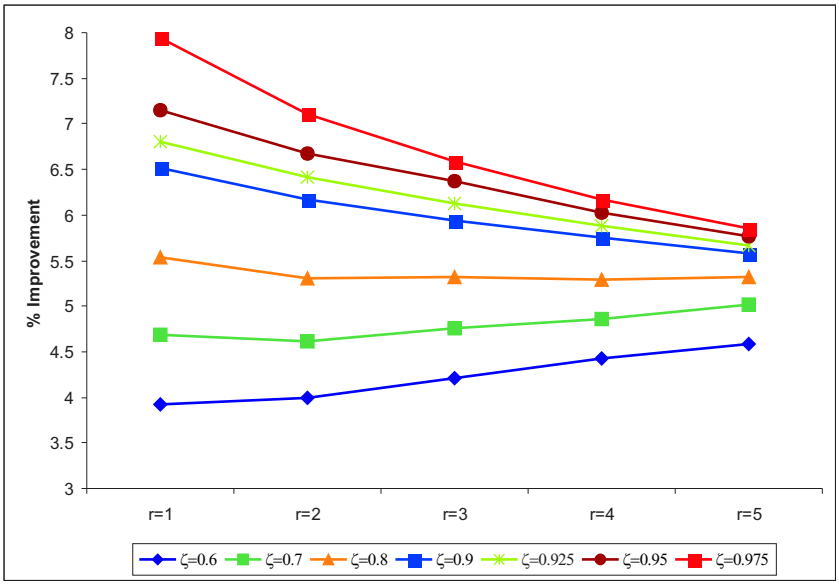



**Figure 12:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for Holding Cost $h = 0.5$

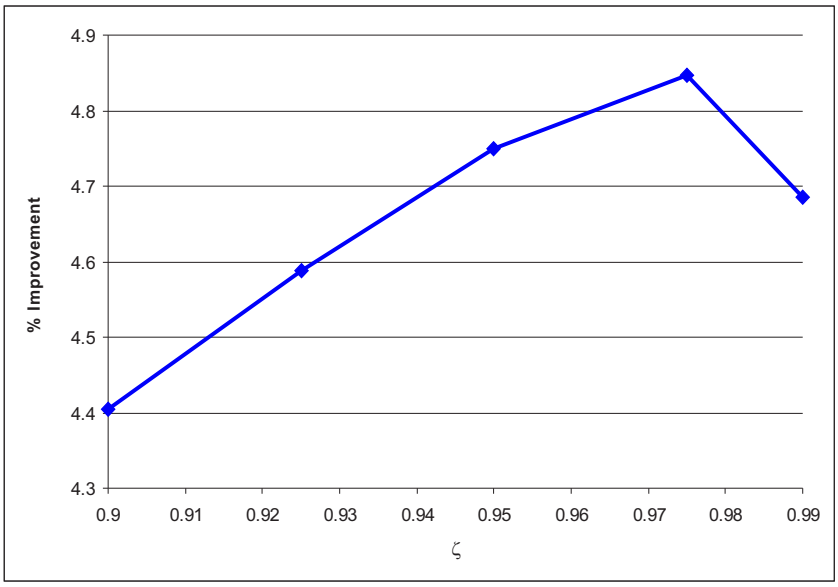



**Figure 13:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for Holding Cost $h = 0.2$ and $r = 2$

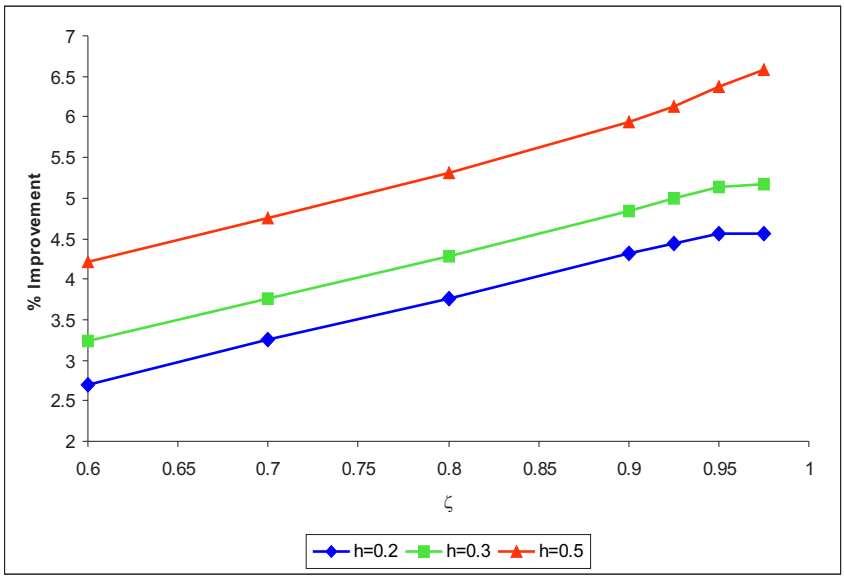



**Figure 14:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for Low Holding Cost Rates and $r = 3$

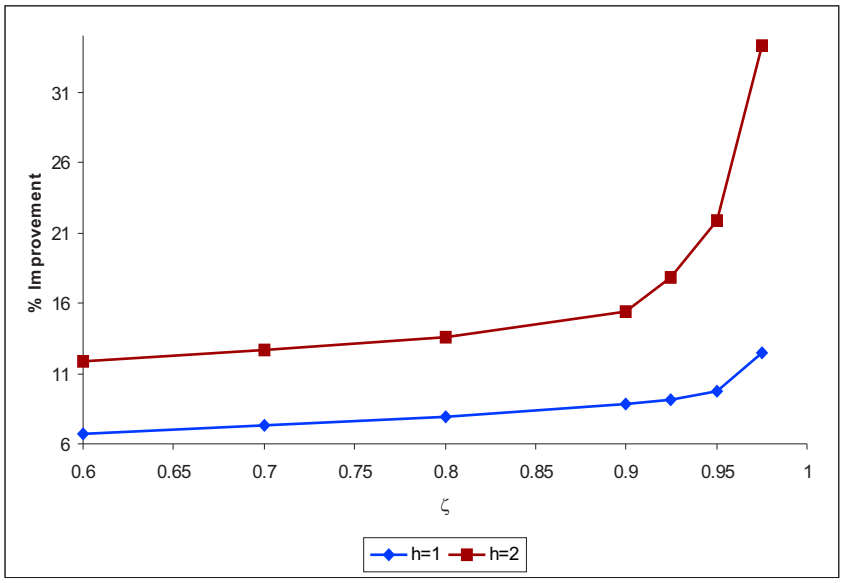



**Figure 15:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for High Holding Cost Rates and $r = 3$

# CHAPTER IV

# BOUNDING THE VALUE OF IMPROVING DEMAND OBSERVABILITY FOR TWO ITEM INVENTORY CONTROL WITH DEMAND SUBSTITUTION AND LOST SALES

## *4.1 Introduction*

In this chapter we develop and analyze an extension of the model considered in Chapter 3. As illustrated earlier, sales data provide censored observations of demand in single product inventory systems. Consider now a scenario with multiple products, and suppose that some of the products are substitutes (imperfect) but that customers only substitute if their preferred product is unavailable. In such cases, sales data for a specific product may underestimate true demand in periods of shortage, but may overestimate in periods of surplus if customers are purchasing the product as a substitute for another unavailable product.

The main objective of this chapter is to understand the value of improved demand observability for inventory control systems with product substitution. We focus on a simple supply chain where a single capacity-constrained production site supplies a single consumption site. The consumption site operates a periodic review inventory control system where orders placed at the end of a period are available at the beginning of the next period. Demand of product 1 at the consumption site is non-stationary and discrete and is assumed to be described by a stationary Markov chain with known transition probabilities. Demand of product 2 at the consumption site is assumed to be stationary and discrete with a known probability distribution. Further, in case of a stockout of product 1, customers may use product 2 as a substitute if it is available, and the conditional distribution of this substitution demand is known. Since product 1 does not substitute for product 2, we call this one way substitution.

We develop a POMDP control model for this system. Sales and inventory levels are

assumed to be perfectly observed. In addition, we assume that we also receive information from a demand sensing program that may help to further identify true demand values.

In this chapter we first present this POMDP model, and develop an algorithm for determining an optimal policy for the expected total discounted profit. Again, since this optimal approach is likely to be intractable, we develop three heuristic algorithms, the third of which is based in a non-standard sufficient statistic that enables relatively easy software implementation. We then analyze two extreme cases of the POMDP model, the completely-observed case and the sales-only case, and use them to determine a bound of the value of improved demand observability in this setting. Finally, we demonstrate the bounding technique for a set of example problem scenarios and investigate the impact of scenario characteristics on the value of improved demand observability. Our numerical results suggest that systems with high levels of substitutability can benefit more from demand observability than systems with lower substitution levels.

## *4.2 Related Literature*

Literature on inventory control with substitution can be roughly categorized into four classes. Research considers both single and multi-period problems, with either two products or multiple ($> 2$) products. Most literature in this area focuses on single period problems with two products. To our knowledge, the two-product multi-period inventory problem with substitution and unobservable demand that we will consider in this dissertation has not been previously considered. Below, we summarize a sample of the research literature on inventory control with demand substitution.

Parlar and Goyal [37] consider a single period inventory model with two products and two way substitution when stockouts occur. Each product demand is assumed to be stochastic and independent with known probability distribution. Further, the proportion of unsatisfied customers that decides to substitute is assumed to be fixed and known for each product. The objective is to maximize expected profit where initial inventory levels are assumed to be zero. Also, the salvage value and lost sales penalty are assumed to be equal to zero. Under these assumptions, they show that the expected profit is a strictly concave function in the

order quantity decision variables for certain values of selling prices and provide necessary and sufficient optimality conditions for these cases.

Pasternack and Drezner [39] study a two product single period inventory system where in the case of a stockout substitution occurs with probability one, but at a different revenue level. They show that the expected profit function is concave and provide formulas for the optimal order quantities. Further, they prove that if the revenue available from substitution of one product for another increases, one will order more of that product and less of the other. For the case of one way substitution, they show that as the transfer revenue increases, the optimal stocking level of the product that can be use as a substitute will increase while the other product's optimal inventory level will decrease. They compare the one way substitution case and the no substitution case and show that the optimal inventory level of the product that serves as a substitute in the one way substitution case will be greater than if substitution were not possible, while the optimal stocking quantity for the good for which substitution occurs will be less than in the case of no substitution. They further show that the *total* optimal stocking levels when substitution is allowed can either be higher or lower than the *total* optimal stocking levels when substitution is not allowed.

Zhand and Chen [58] study the optimal joint replenishment policy for a single period two product inventory system with stochastic demands and one way substitution. Assuming a fixed setup cost, the objective is to maximize the expected profit. They prove that the profit function is convex and supermodular and provide the structure of the optimal joint replenishment policy.

Nagarajan and Rajagopalan [34] consider an inventory system where total demand is fixed and known, but demand of each product is random. They first study the single period two product case with negative correlated demand where substitution occurs at a fixed proportion, assuming the same cost parameters for both products. For certain levels of substitution, closed-form formulas for the optimal inventory levels are provided; they further prove that the base stock level of one product is independent of the inventory level of the other, referring to this case as a "partially decoupled" policy. They present the finite horizon case and prove that the profit function is concave. For this case and for some

levels of substitution, "partially decoupled" policies are also optimal and the order-up-to level is shown to be monotonically increasing in the number of periods until the end of the horizon. They also consider the infinite horizon case, where they show that for certain levels of substitution the optimal stationary policy is a base stock level policy and partially decoupled. Next, they consider the $N$-product single period model with identical cost parameters for all products. Each customer has a first choice product and a fixed proportion of unsatisfied customers will take any other product as a substitute (*i.e.*, indifferent second choice). The profit function is proven to be concave and for certain levels of substitution a close form for the optimal stock level was provided. They consider a "duopoly" model, in which they examine two retailers with identical cost structure competing in a market where total demand is $D$. They characterize the equilibrium decisions of the two players in a single period model and show that under certain conditions the equilibrium quantities are such that each player ignores the strategy of its opponent. Further, they extend the model to a finite horizon multi-period version and show that under certain circumstances the equilibrium quantities are obtained by solving a multi-period single product inventory model for each player.

Ernst and Kouvelis [14] consider a single period inventory system with two products that are sold either independently or in a package containing one of each. There is no substitution between individual products, but there is substitution in any direction between individual products and the package. A fixed and known proportion of unsatisfied customers of an individual product will accept the package as a substitute, and a fixed and known proportion of customers that have as first option the package will substitute for either one or both individual products. They show that the expected profit function is concave and continuously differentiable, sufficient and necessary optimality conditions and a numerical search procedure for obtaining optimal stocking levels are provided. An extensive computational study is performed to compare optimal stocking policies with simple independent newsboy policies, and to determine effects on demand correlation and fraction of substitution. The main conclusions obtained from this computational study are: (1) the use of independent newsboy policies leads to suboptimal stocking policies, (2) positive correlation

of demand tends to increase the stocking level of the package, (3) demand correlation results in higher profitability of inventory systems as compared with uncorrelated systems and (4) the stronger the substitution pattern the higher the optimal stocking levels of the package.

Bassok et al. [5] present a single period multi-product inventory system with random demand. They consider $N$ products and $N$ demand classes with full downward demand substitution (*i.e.*, excess demand for class $i$ can be satisfied using product $j$ where $i \geq j$). They assume a substitution cost that is proportional to the quantity substituted and propose a two stage model where first they make ordering decisions and then allocation decisions (*i.e.*, allocation of excess inventory of class $k$ among unsatisfied demands of classes $k+1, k+2, \cdots N$). They prove that the profit function is concave and submodular and that the optimal policy is an order-up-to level policy. A computational study is performed for the two product case and profit gain between the optimal solution with the profits obtained when order points are derived using the standard newsvendor model is compared. They conclude that most gains accrue in a problem with high salvage value of products, high demand variability, low profit margins and similarity of products in terms of prices and costs.

Smith and Agrawal [46] consider a single period multi-product inventory policy that maximizes the expected profit subject to a variety of constraints, which may include floor space, budget and assortment size; the policy specifies both the items to be stocked and the initial inventory level of each item. Demand of each item is random and in case of a stockout, a customer selects a substitute randomly with known probabilitys from a choice set. They determine order quantities of each product based on a predefined service level. The focus of the paper is the estimation of the demand of each item taking into consideration that demand is affected by substitution.

Netessine and Rudi [36] consider a multi-product single period inventory system where unsatisfied demand of a product flow to other products in deterministic proportions. They compare centralized inventory management with inventory management under competition (*i.e.*, decentralized system where each product is managed by a an independent decision maker maximizing the expected profit generated by this specific product while interacting

strategically with other decision makers) and show that there are situations in which the optimal stocking quantity of an item is higher in the centralized system than in the decentralized system. However, there exists at least one product for which the optimal stocking quantity is lower in the centralized system than in the decentralized system. They also show that if costs and revenues are symmetric among firms, demands are i.i.d and customers are equally likely to switch to any product, the stocking quantities of the decentralized systems are at least as high as the stocking quantities of the centralized system.

Rao at el. [41] present a single period multi-product inventory model with downward substitution and set up costs. They formulate the problem as a two stage integer stochastic program with recourse where the first stage variables determine which products to produce and production quantities, and the second stage variables determine the allocation of products to satisfy the realized demand. An efficient solution technique is presented.

Literature on multi-period and multi-product inventory systems with substitution is scarce. McGillivray and Silver [31] study a $(R, S)$ inventory control system for $N$ products with identical cost parameters where demands are independent and normally distributed. A fixed and known proportion $a_{ij}$ of unsatisfied customers whose first choice was product $i$ will take product $j$ as a substitute. To determine potential savings due to substitution, they consider two extreme cases: (i) no substitution ($a_{ij} = 0$ for all $i, j$) and (ii) complete substitution ($a_{ij} = 1$ for all $i, j$); based on the comparison of these cases, they conclude that accounting for substitution can lead to significant savings, especially when the number of items is high. Simulation is used to solve the general case (*i.e.*, $0 < a_{ij} < 1$) ) for two items, results obtained suggest that when one of the substitution probabilities is close to 1, the optimal stocking rule is substantially different from the case where the items are treated independently.

## *4.3 Two Product Inventory Control with One Way Demand Substitution*

Consider a two product supply chain system with one way demand substitution in which a single decision maker selects replenishment quantities at each of a discrete, predefined and finite set of decision epochs. The objective is to maximize total expected discounted

profit over the problem horizon. Just prior to each decision epoch, the decision maker observes current inventory levels and the demands that have occurred since the last decision epoch. We assume that the inventory levels are completely observed but that the demand observations may be noise corrupted. Selection of the replenishment quantities at the current epoch are based on all past and present inventory and demand observations and all past ordering decisions. We assume that the ordered quantities are received immediately after order. Furthermore, demand of product type 1 is assumed to be described by a control-independent (exogenous) Markov chain, and demand of product type 2 is assumed to be independent and identically distributed with a known probability distribution function. Substitution may occur if there is a stockout of product type 1. Specifically, each unsatisfied unit demand of product type 1 may be substituted with an available unit of product type 2 with probability $\alpha$.

More precisely, let $x_i(t)$ be the completely observed inventory level of product type $i$ at time (or decision epoch) $t$, just prior to the selection of replenishment decisions $a_1(t)$ and $a_2(t)$, $i = 1, 2$. Let $d_i(t)$ be the primary demand of product type $i$ realized between time $t-1$ and time $t$ for $i = 1, 2$. Let $d_2'(t)$ be the secondary demand of product type 2 generated by substitution realized between time $t-1$ and time $t$. Note that $(d_2'(t)|d_1(t), x_1(t), a_1(t))$ follows a Binomial distribution with parameters $[d_1(t) - x_1(t) - a_1(t)]^+$ and $\alpha$. We assume replenishment decisions are made at each $t \in \{0, 1, \cdots, T-1\}$, where $T < \infty$. Thus, the planning horizon is finite. No backlogging is permitted, and hence

$$x_1(t+1) = \max\{0, x_1(t) + a_1(t) - d_1(t+1)\} \tag{9}$$

and

$$x_2(t+1) = \max\{0, x_2(t) + a_2(t) - d_2(t+1) - d_2'(t+1)\}. \tag{10}$$

Let $z_1(t)$ be the noise-corrupted observation of type 1 demand that is available just prior to the selection of replenishment decisions, and assume probabilities of the form $P(z, j|i) = P(z_1(t+1) = z, d_1(t+1) = j|d_1(t) = i)$ are given. We note that $P(z, j|i) = P(z|j, i)P(j|i)$ where $P(j|i) = \sum_z P(z, j|i) = P(d_1(t+1) = j|d_1(t) = i)$ and

$$P(z|j,i) = \frac{(z,j|i)}{P(j|i)} = P(z_1(t+1) = z|d_1(t+1) = j, d_1(t) = i),$$

assuming $P(j|i) \neq 0$. The probabilities $P(j|i)$ and $P(z|j,i)$ are referred to as *transition* and *observation* probabilities, respectively. Further, the known demand distribution function for product type 2 is given by $\overline{P}(i) = P(d_2(t) = i)$. We remark that given that demand of product 2 is assumed to be *i.i.d.*, there is no need to attempt to improve its observability since its distribution in a given period is always known. We assume that $D_i$ is the maximum demand per period for product $i$, and hence $d_i(t) \in \{0, 1, \cdots, D_i\}$. Similarly, we assume that $z_1(t) \in \{0, 1, \cdots, D_1\}$.

We will have particular interest in two special cases of the type 1 observation probabilities. We note that

$$z_1(t) = d_1(t)$$

*w.p.*1 for all $t$ is equivalent to

$$P(z|j,i) = \begin{cases} 1 & \text{if } z = j \\ 0 & \text{otherwise} \end{cases} \qquad \forall \quad i$$

For observation processes with this characteristic, we say that type 1 demand is completely (or perfectly) observed by the observation process $\{z_1(t), t = 1, 2, \cdots\}$. Alternatively, if $P(z|j,i)$ is independent of $i$ and $j$, then the observation process provides no information about demand of product type 1, and hence we say that demand type 1 is completely unobserved by the observation process. We remark that when the observation process provides no information about the demand process, information about the demand process of product type 1 can be inferred from the inventory processes, or equivalently, from sales data of both products, which will be described in section 4.4.

Selection of $a(t) = \{a_1(t), a_2(t)\}$ are made with knowledge of the information set at time $t$, $\mathcal{H}(t)$, where $\mathcal{H}(t) = \Big\{z_1(t), \cdots, z_1(1), x_1(t), \cdots, x_1(0), x_2(t), \cdots, x_2(0), a_1(t-1), \cdots, a_1(0), a_2(t-1), \cdots, a_2(0), \xi^1(0)\Big\}$, $\xi^1(0) = \Big\{\xi_i^1(0)\Big\}$, and $\xi_i^1(0) = P(d_1(0) = i)$. (Note that $\xi^1(0) \in \Xi^1 = \Big\{\xi^1 \geq 0 : \sum_{i=0}^{D_1} \xi_i^1 = 1\Big\}$). Hence, the order at epoch $t$, $a(t)$, is

allowed to depend on all past and present (possibly corrupted) observations of demand type 1, all past and present inventory levels, all former replenishment orders, and an *a priori* type 1 demand information.

Let $p_i$, $\bar{p}_i$, $c_i$, and $h_i$ be the per unit selling price, salvage value, ordering cost, and per period inventory holding cost, of product type $i$. We assume that holding costs accrued from $t$ to $t+1$ are determined on the basis of inventory levels at time $t$, $x_1(t)$ and $x_2(t)$.

A policy $\pi$ is a rule that determines actions $a_1$ and $a_2$ on the basis of the information currently available. Thus, $[a_1(t), a_2(t)] = \pi(t, \mathcal{H}(t))$ for all $t \in \{0, 1, \cdots, T-1\}$.

The *Inventory Replenishment Problem with Demand Substitution (IRPDS)* objective is to find a policy that maximizes the following criterion with respect to all policies:

$$\mathbf{E}^{\pi}_{\xi^1(0)} = \left\{ \sum_{t=0}^{T-1} \beta^t r\left[s(t), a(t)\right] + \beta^T \bar{r}\left[s(T)\right] \right\}, \tag{11}$$

where $a(t) = [a_1(t), a_2(t)]$, $s(t) = [x_1(t), x_2(t), d_1(t)]$ and $\mathbf{E}^{\pi}_{\xi^1(0)}$ is the expectation operator conditioned on $\xi^1(0)$ and use of policy $\pi$, $\beta$ is the discount factor, and where $r\left[s(t), a(t)\right] = -h_1 x_1(t) - h_2 x_2(t) - c_1 a_1(t) - c_2 a_2(t) + p_1 \mathbf{E}\Big\{ \min\{d_1(t+1), x_1(t) + a_1(t)\} \Big\} + p_2 \mathbf{E}\Big\{ \min\{d_2(t+1) + d_2'(t+1), x_2(t) + a_2(t)\} \Big\}$ and $\bar{r}\left[s(T)\right] = \bar{p}_1 x(T) + \bar{p}_2 x(T)$. Note that $\min\{d_1(t+1), x_1(t) + a_1(t)\}$ and $\min\{d_2(t+1) + d_2'(t+1), x_2(t) + a_2(t)\}$ represent sales of product 1 and 2 between $t$ and $t+1$ respectively.

We observe that given that there are no fixed ordering costs and no replenishment lead times in the *IRPDS* setting, an optimal policy will always select values of $a(t)$ such that $a_1(t) \leq D_1 - x_1(t)$, $a_2(t) \leq D_1 + D_2 - x_2(t)$ and $a_1(t) + a_2(t) \leq D_1 + D_2 - x_1(t) - x_2(t)$ for all $t \in \{0, \cdots, T-1\}$

### *4.4 Preliminary Results*

The following observations result from equations (9) and (10):

(i) If $x_1(t+1) > 0$, then $d_1(t+1) = x_1(t) + a_1(t) - x_1(t+1)$ and $d_2'(t+1) = 0$. Hence, $d_1(t+1)$ is completely observed and there is no substitution between time $t$ and $t+1$.

(ii) If $x_1(t+1) = 0$ and $x_2(t+1) = \bar{x}_2 > 0$, given $d_2(t+1) \leq D_2$ then at least $[x_2(t) + a_2(t) - \bar{x}_2 - D_2]^+$ type 1 customers should buy product type 2 as a substitute. Therefore,

$d_1(t+1) \geq x_1(t) + a_1(t) + [x_2(t) + a_2(t) - \bar{x}_2 - D_2]^+$ and $[x_2(t) + a_2(t) - \bar{x}_2 - D_2]^+ \leq d_2'(t+1) \leq d_1(t+1) - a_1(t) - x_1(t)$. Additionally, $x_2(t+1) = \bar{x}_2 > 0$ implies no stockouts of product type 2. Therefore $d_2(t+1) + d_2'(t+1) = x_2(t) + a_2(t) - \bar{x}_2$.

(iii) If $x_1(t+1) = 0$ and $x_2(t+1) = 0$, given $d_2(t+1) \leq D_2$ then at least $[x_2(t)+a_2(t)-D_2]^+$ type 1 customers should buy product type 2 as a substitute. Therefore, $d_1(t+1) \geq x_1(t) + a_1(t) + [x_2(t) + a_2(t) - D_2]^+$ and $[x_2(t) + a_2(t) - D_2]^+ \leq d_2'(t+1) \leq d_1(t+1) - x_1(t) - a_1(t)$. In this case, as opposed to case (ii), there could be stockouts of product type 2. Therefore, $d_2(t+1) + d_2'(t+1) \geq x_2(t) + a_2(t)$.

The above observations imply that in case (i) there is perfect demand observability, while in cases (ii) and (iii) only partial demand observability is available. However, in case (ii) more information about demand is provided than in case (iii).

It follows from Smallwood and Sondik [45] that $(x_1(t), x_2(t), \xi^1(t))$ represents a sufficient statistic for the *IRPDS*, where $\xi^1(t) = \{\xi_i^1(t)\} \in \Xi^1$ and $\xi_i^1(t) = P(d_1(t) = i|\mathcal{H}(t))$. This fact, coupled with the above observations, imply that there are three general states of interest:

(i) $(x_1, x_2, e_i)$ for $x_1 > 0$ and for any $x_2$, where the $j^{th}$ element of the vector $e_i$ is 1 if $i = j$ and 0 otherwise.

(ii) $(0, x_2, \xi^1)$ for $x_2 > 0$ and for any probability mass vector $\xi^1$ for the type 1 demand state.

(iii) $(0, 0, \xi^1)$ for any probability mass vector $\xi^1$ for the type 1 demand state.

For $\overline{x} > 0$ let

$$\tilde{\sigma}^1(z, (x_1, x_2, \xi), a_1, a_2, \overline{x}) = \sum_{j \geq \widetilde{\mathbf{J}}_\mathbf{1}} \sum_{l=\widetilde{\mathbf{L}}_\mathbf{1}}^{\widetilde{\mathbf{L}}_\mathbf{2}} \overline{P}(x_2 + a_2 - \overline{x} - l) B(j, l, \alpha) \sum_i \xi_i P(z, j|i),$$

where $\widetilde{\mathbf{J}}_\mathbf{1} = x_1 + a_1 + [x_2 + a_2 - \overline{x} - D_2]^+$, $\widetilde{\mathbf{L}}_\mathbf{1} = [x_2 + a_2 - \overline{x} - D_2]^+$ and $\widetilde{\mathbf{L}}_\mathbf{2} = \min\{j - a_1 - x_1, x_2 + a_2 - \overline{x}\}$. If $j \geq \widetilde{\mathbf{J}}_\mathbf{1}$, let

$$\tilde{\lambda}_j^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x}) = \frac{\sum_{l=\widetilde{\mathbf{L_1}}}^{\widetilde{\mathbf{L_2}}} \overline{P}(x_2+a_2-\overline{x}-l)B(j,l,\alpha)\sum_i \xi_i P(z,j|i)}{\tilde{\sigma}^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x})}.$$

Otherwise, $\tilde{\lambda}_j^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x}) = 0$, where $\tilde{\sigma}^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x}) \neq 0$, $\tilde{\lambda}^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x}) = \{\tilde{\lambda}_j^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x})\}$ and $B(j,l,\alpha) = \binom{j}{l}\alpha^l(1-\alpha)^{j-l}$.

Also, let

$$\tilde{\sigma}^2(z,(x_1,x_2,\xi),a_1,a_2) = \sum_{j\geq \widetilde{\mathbf{J_2}}} \sum_{l=\widetilde{\mathbf{L_3}}}^{\widetilde{\mathbf{L_4}}} \sum_{m\geq [x_2+a_2-l]^+} \overline{P}(m)B(j,l,\alpha)\sum_i \xi_i P(z,j|i),$$

where $\widetilde{\mathbf{J_2}} = x_1+a_1+[x_2+a_2-D_2]^+$, $\widetilde{\mathbf{L_3}} = [x_2+a_2-D_2]^+$ and $\widetilde{\mathbf{L_4}} = j-x_1-a_1$. If $j \geq \widetilde{\mathbf{J_2}}$, let

$$\tilde{\lambda}_j^2(z,(x_1,x_2,\xi),a_1,a_2) = \frac{\sum_{l=\widetilde{\mathbf{L_3}}}^{\widetilde{\mathbf{L_4}}} \sum_{m\geq [x_2+a_2-l]^+} \overline{P}(m)B(j,l,\alpha)\sum_i \xi_i P(z,j|i)}{\tilde{\sigma}^2(z,(x_1,x_2,\xi),a_1,a_2)}.$$

Otherwise, $\tilde{\lambda}_j^2(z,(x_1,x_2,\xi),a_1,a_2) = 0$, where $\tilde{\sigma}^2(z,(x_1,x_2,\xi),a_1,a_2) \neq 0$ and $\tilde{\lambda}^2(z,(x_1,x_2,\xi),a_1,a_2) = \{\tilde{\lambda}_j^2(z,(x_1,x_2,\xi),a_1,a_2)\}$.

Let $\psi_t(a_1,a_2,x_1,x_2,\xi) = \{a_1(t)=a_1, a_2(t)=a_2, x_1(t)=x_1, x_2(t)=x_2, \xi^1(t)=\xi\}$, and note that:

- $\tilde{\sigma}^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x}) = P\left(\begin{array}{c|c} x_1(t+1)=0, & \psi_t(a_1,a_2,x_1,x_2,\xi), \\ z_1(t+1)=z & x_2(t+1)=\overline{x} \end{array}\right)$
- $\tilde{\sigma}^2(z,(x_1,x_2,\xi),a_1,a_2,\overline{x}) = P\left(\begin{array}{c|c} x_1(t+1)=0, & \psi_t(a_1,a_2,x_1,x_2,\xi), \\ z_1(t+1)=z & x_2(t+1)=0 \end{array}\right)$

- $\tilde{\lambda}^1_j(z,(x_1,x_2,\xi),a_1,a_2,\overline{x}) = P\left(d_1(t+1)=j \,\middle|\, \begin{array}{c}\psi_t(a_1,a_2,x_1,x_2,\xi),x_1(t+1)=0,\\ x_2(t+1)=\overline{x},z_1(t+1)=z\end{array}\right)$

- $\tilde{\lambda}^2_j(z,(x_1,x_2,\xi),a_1,a_2) = P\left(d_1(t+1)=j \,\middle|\, \begin{array}{c}\psi_t(a_1,a_2,x_1,x_2,\xi),x_1(t+1)=0,\\ x_2(t+1)=0,z_1(t+1)=z\end{array}\right)$.

Thus, assuming $(x_1(t),x_2(t),\xi^1(t)) = (x_1,x_2,\xi)$:

(i) If $x_1(t+1) > 0$ then $\xi^1(t+1) = e_i$, where $d_1(t+1) = x_1(t) + a_1(t) - x_1(t+1) = i$.

(ii) If $x_1(t+1) = 0$ and $x_2(t+1) = \overline{x} > 0$ then $\xi^1(t+1) = \tilde{\lambda}^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x})$ with probability $\tilde{\sigma}^1(z,(x_1,x_2,\xi),a_1,a_2,\overline{x})$ where $a_1(t) = a_1$, $a_2(t) = a_2$ and $z_1(t+1) = z$.

(iii) If $x_1(t+1) = 0$ and $x_2(t+1) = 0$ then $\xi^1(t+1) = \tilde{\lambda}^2(z,(x_1,x_2,\xi),a_1,a_2)$ with probability $\tilde{\sigma}^2(z,(x_1,x_2,\xi),a_1,a_2)$ where $a_1(t) = a_1$, $a_2(t) = a_2$ and $z_1(t+1) = z$.

We now present optimality equations for the three cases. In all cases, $v_T(x_1,x_2,\xi) = \bar{p}_1x_1 + \bar{p}_2x_2$ for all $\xi$.

If $x_1 > 0$, then:

$$
\begin{aligned}
v_t(x_1,x_2,e_i) = \max_{\substack{a_1\geq 0\\ a_2\geq 0}} \Bigg\{ & -\sum_{i=1}^{2}(h_ix_i + c_ia_i) + p_1\sum_{k=0}^{D_1}\min\{k,x_1+a_1\}P(k|i) + \\
& p_2\left[\sum_{k=0}^{D_1+D_2}\min\{k,x_2+a_2\}\left(\sum_{r\leq a_1+x_1}\overline{P}(k)P(r|i) + \right.\right. \\
& \left.\left.\sum_{r>a_1+x_1}\sum_{\substack{l=k-r+\\ x_1+a_1}}^{k} B(r-x_1-a_1,k-l,\alpha)\overline{P}(l)P(r|i)\right)\right] + \\
& \beta\sum_{j<x_1+a_1}\sum_{k}\overline{P}(k)P(j|i)v_{t+1}(x_1+a_1-j,[x_2+a_2-k]^+,e_j) + \\
& \beta\sum_{z}\sum_{x=1}^{x_2+a_2}\tilde{\sigma}^1(z,(x_1,x_2,e_i),a_1,a_2,x)v_{t+1}(0,x,\tilde{\lambda}^1(z,(x_1,x_2,e_i),a_1,a_2,x)) + \\
& \beta\sum_{z}\tilde{\sigma}^2(z,(x_1,x_2,e_i),a_1,a_2)v_{t+1}(0,0,\tilde{\lambda}^2(z,(x_1,x_2,e_i),a_1,a_2))\Bigg\}.
\end{aligned}
$$

If $x_1 = 0$ and $x_2 > 0$, then:

$$
\begin{aligned}
v_t(0,x_2,\xi) \;=\; \max_{\substack{a_1\geq 0\\ a_2\geq 0}} \Bigg\{ & -h_2 x_2 - \sum_{i=1}^{2}(c_i a_i) + p_1 \sum_{k=0}^{D_1} \min\{k,a_1\} \sum_i \xi_i P(k|i) + \\
& p_2 \left[ \sum_{k=0}^{D_1+D_2} \min\{k, x_2+a_2\} \left( \sum_{r\leq a_1} \overline{P}(k) \sum_i \xi_i P(r|i) + \right.\right. \\
& \left.\left. \sum_{r>a_1} \sum_{l=k-r+a_1}^{k} B(r-a_1, k-l, \alpha) \overline{P}(l) \sum_i \xi_i P(r|i) \right) \right] + \\
& \beta \sum_{j<a_1} \sum_k \overline{P}(k) \sum_i \xi_i P(j|i) v_{t+1}(a_1 - j, [x_2 + a_2 - k]^+, e_j) + \\
& \beta \sum_z \sum_{x=1}^{x_2+a_2} \tilde{\sigma}^1(z,(0,x_2,\xi),a_1,a_2,x) v_{t+1}(0,x,\tilde{\lambda}^1(z,(0,x_2,\xi),a_1,a_2,x)) + \\
& \beta \sum_z \tilde{\sigma}^2(z,(0,x_2,\xi),a_1,a_2) v_{t+1}(0,0,\tilde{\lambda}^2(z,(0,x_2,\xi),a_1,a_2)) \Bigg\}.
\end{aligned}
$$

If $x_1 = 0$ and $x_2 = 0$, then:

$$
\begin{aligned}
v_t(0,0,\xi) \;=\; \max_{\substack{a_1\geq 0\\ a_2\geq 0}} \Bigg\{ & -\sum_{i=1}^{2}(c_i a_i) + p_1 \sum_{k=0}^{D_1} \min\{k,a_1\} \sum_i \xi_i P(k|i) + \\
& p_2 \left[ \sum_{k=0}^{D_1+D_2} \min\{k, a_2\} \left( \sum_{r\leq a_1} \overline{P}(k) \sum_i \xi_i P(r|i) + \right.\right. \\
& \left.\left. \sum_{r>a_1} \sum_{l=k-r+a_1}^{k} B(r-a_1, k-l, \alpha) \overline{P}(l) \sum_i \xi_i P(r|i) \right) \right] + \\
& \beta \sum_{j<a_1} \sum_k \overline{P}(k) \sum_i \xi_i P(j|i) v_{t+1}(a_1 - j, [a_2 - k]^+, e_j) + \\
& \beta \sum_z \sum_{x=1}^{a_2} \tilde{\sigma}^1(z,(0,0,\xi),a_1,a_2,x) v_{t+1}(0,x,\tilde{\lambda}^1(z,(0,0,\xi),a_1,a_2,x)) + \\
& \beta \sum_z \tilde{\sigma}^2(z,(0,0,\xi),a_1,a_2) v_{t+1}(0,0,\tilde{\lambda}^2(z,(0,0,\xi),a_1,a_2)) \Bigg\}.
\end{aligned}
$$

For $y_i = x_i + a_i$ define:

$$
\begin{aligned}
\sigma^1(z,\xi,y_1,y_2,x) &= \tilde{\sigma}^1(z,(x_1,x_2,\xi),a_1,a_2,x)\\
\sigma^2(z,\xi,y_1,y_2) &= \tilde{\sigma}^2(z,(x_1,x_2,\xi),a_1,a_2)\\
\lambda^1(z,\xi,y_1,y_2,x) &= \tilde{\lambda}^1(z,(x_1,x_2,\xi),a_1,a_2,x)\\
\lambda^2(z,\xi,y_1,y_2) &= \tilde{\lambda}^2(z,(x_1,x_2,\xi),a_1,a_2)\\
L_1(\xi,y_1) &= p_1\sum_i \xi_i \sum_{k=0}^{D_1} \min\{k,y_1\}P(k|i)\\
L_2(\xi,y_1,y_2) &= p_2\sum_i \xi_i \sum_{k=0}^{D_1+D_2} \min\{k,y_2\}\left(\sum_{r\le y_1}\overline{P}(k)P(r|i)\right.\\
&\quad \left.+\sum_{r>y_1}\sum_{l=k-r+y_1}^{k} B(r-y_1,k-l,\alpha)\overline{P}(l)P(r|i)\right).
\end{aligned}
$$

Note that $L_1(\xi,y_1)=\sum_i \xi_i L_1(e_i,y_1)$ and $L_2(\xi,y_1,y_2)=\sum_i \xi_i L_2(e_i,y_1,y_2)$. Also let:

$$
\begin{aligned}
h(\xi,y_1,y_2,v) &= -\sum_i c_i y_i + L_1(\xi,y_1) + L_2(\xi,y_1,y_2)\\
&\quad +\beta\sum_{j<y_1}\sum_i \xi_i \sum_k \overline{P}(k)P(j|i)v(y_1-j,[y_2-k]^+,e_j)\\
&\quad +\beta\sum_z \sum_{x=1}^{y_2} \sigma^1(z,\xi,y_1,y_2,x)v(0,x,\lambda^1(z,\xi,y_1,y_2,x))\\
&\quad +\beta\sum_z \sigma^2(z,\xi,y_1,y_2)v(0,0,\lambda^2(z,\xi,y_1,y_2))
\end{aligned}
$$

and

$$
[Hv](x_1,x_2,\xi)=\sum_i (c_i-h_i)x_i + \max_{\substack{y_1\ge x_1\\ y_2\ge x_2}} h(\xi,y_1,y_2,v).
$$

Then the optimality equation is $v_t = Hv_{t+1}$, where $v_T(x_1,x_2,\xi)=\bar{p}_1x_1+\bar{p}_2x_2$ for all $\xi$ and:

(i) $v_t(x_1,x_2,\xi)$ is the optimal expected reward to be accrued from $t$ until $T$ given $x_1(t)=x_1$, $x_2(t)=x_2$ and $\xi^1(t)=\xi$.

(ii) A set of actions that causes the maximum in the optimality equation to be attained are optimal actions for the concomitant state.

## *4.5 Numerical Algorithms*

We now present two general approaches for determining optimal policies and three heuristic algorithms for determining a sub-optimal policy for the *IRPDS* and the expected value accrued over the planning horizon. The approaches differ on the basis of the sufficient statistic used.

### 4.5.1 Approach 1

The first optimal algorithm uses $(x_1(t), x_2(t), \xi(t))$ as a sufficient statistic for $\mathcal{H}(t)$ and takes advantage of the fact that for each $t$ and $x_2$, there is a finite set of vectors, $\Gamma_t(x_2)$, such that $v_t(0, x_2, \xi) = \max\{\xi\gamma : \gamma \in \Gamma_t(x_2)\}$; that is, $v_t(0, x_2, \xi)$ is piecewise linear convex in $\xi$ for all $x_2$ for finite $T$. Thus, although the set of all probability mass vectors $\xi$ is uncountably infinite, $v_t(0, x_2, .)$ has a finite representation.

$\Gamma_t(x_2)$ can be constructed from $\{\Gamma_{t+1}(0), \cdots, \Gamma_{t+1}(D_1 + D_2)\}$ as follows. Note $v_T(0, 0, \xi) = 0$ for all $\xi$ and $v_T(0, x_2, \xi) = \bar{p}_2 x_2$ for all $x_2 > 0$ and for all $\xi$. Thus, $\Gamma_T(0) = \{0\}$ and $\Gamma_T(x_2) = \{\bar{p}_2 x_2 \mathbf{1}\}$ for all $x_2$, where $\mathbf{1}$ is the $(D_1 + 1)$-dimensional vector of ones. Then,

$$
\begin{aligned}
v_t(0,x_2,\xi) &= (c_2-h_2)x_2+\max_{\substack{y_1\geq 0\\ y_2\geq x_2}}\Bigg\{-\sum_j c_j y_j + L_1(\xi,y_1)+L_2(\xi,y_1,y_2)+\\
&\quad \beta\sum_{j<y_1}\sum_i \xi_i \sum_k \overline{P}(k)P(j|i)v(y_1-j,[y_2-k]^+,e_j)+\\
&\quad \beta\sum_z\sum_{x=1}^{y_2}\sigma^1(z,\xi,y_1,y_2,x)\max\{\lambda^1(z,\xi,y_1,y_2,x)\gamma(x):\gamma(x)\in\Gamma_{t+1}(x)\}+\\
&\quad \beta\sum_z \sigma^2(z,\xi,y_1,y_2)\max\{\lambda^2(z,\xi,y_1,y_2)\gamma(0):\gamma(0)\in\Gamma_{t+1}(0)\}\Bigg\}\\
&= \max_{\substack{y_1\geq 0\\ y_2\geq x_2}}\max_{\gamma(0)^0}\cdots\max_{\gamma(0)^Z}\cdots\max_{\gamma(y_2)^0}\cdots\max_{\gamma(y_2)^Z}\Bigg\{(c_2-h_2)x_2-\sum_j c_j y_j+\\
&\quad L_1(\xi,y_1)+L_2(\xi,y_1,y_2)+\\
&\quad \beta\sum_{j<y_1}\sum_i \xi_i \sum_k \overline{P}(k)P(j|i)v(y_1-j,[y_2-k]^+,e_j)+\\
&\quad \beta\sum_z\sum_{x=1}^{y_2}\sigma^1(z,\xi,y_1,y_2,x)\lambda^1(z,\xi,y_1,y_2,x)\gamma(x)^z+\\
&\quad \beta\sum_z \sigma^2(z,\xi,y_1,y_2)\lambda^2(z,\xi,y_1,y_2)\gamma(0)^z\Bigg\}.
\end{aligned}
$$

It follows that:

$$
\sigma^1(z,\xi,y_1,y_2,x)\lambda^1(z,\xi,y_1,y_2,x)\gamma(x)^z=\sum_{j\geq \mathbf{J_1}}\sum_{l=\overline{\mathbf{L}}_\mathbf{1}}^{\overline{\mathbf{L}}_\mathbf{2}}\overline{P}(y_2-x-l)B(j,l,\alpha)\sum_i \xi_i P(z,j|i)\gamma_j(x)^z
$$

and

$$
\sigma^2(z,\xi,y_1,y_2)\lambda^2(z,\xi,y_1,y_2)\gamma(0)^z=\sum_{j\geq \mathbf{J_2}}\sum_{l=\overline{\mathbf{L}}_\mathbf{3}}^{\overline{\mathbf{L}}_\mathbf{4}}\sum_{m=y_2-l}^{D_2}\overline{P}(m)B(j,l,\alpha)\sum_i \xi_i P(z,j|i)\gamma_j(0)^z
$$

where $\mathbf{J_1} = y_1+[y_2-x-D_2]^+$, $\mathbf{J_2} = y_1+[y_2-D_2]^+$, $\overline{\mathbf{L}}_\mathbf{1} = [y_2-x-D_2]^+$,$\overline{\mathbf{L}}_\mathbf{2} = \min\{j-y_1, y_2-x\}$, $\overline{\mathbf{L}}_\mathbf{3} = [y_2-D_2]^+$ and $\overline{\mathbf{L}}_\mathbf{4} = j-y_1$. Hence,

$$
\begin{aligned}
v_t(0,x_2,\xi) \;=\; & \max_{\substack{y_1\geq 0\\ y_2\geq x_2}} \max_{\gamma(0)^0}\cdots\max_{\gamma(0)^Z}\cdots\max_{\gamma(y_2)^0}\cdots\max_{\gamma(y_2)^Z}\left\{\sum_i \xi_i\left[(c_2-h_2)x_2-\sum_j c_j y_j+\right.\right.\\
& L_1(e_i,y_1)+L_2(e_i,y_1,y_2)+\sum_{j<y_1}\sum_k \overline{P}(k)P(j|i)v(y_1-j,[y_2-k]^+,e_j)+\\
& \beta\sum_z\sum_{x=1}^{y_2}\sum_{j\geq \mathbf{J_1}}\sum_{l=\overline{\mathbf{L}}_{\mathbf{1}}}^{\overline{\mathbf{L}}_{\mathbf{2}}}\overline{P}(y_2-x-l)B(j,l,\alpha)P(z,j|i)\gamma_j(x)^z+\\
& \left.\left.\beta\sum_z\sum_{j\geq \mathbf{J_2}}\sum_{l=\overline{\mathbf{L}}_{\mathbf{3}}}^{\overline{\mathbf{L}}_{\mathbf{4}}}\sum_{m=y_2-l}^{D_2}\overline{P}(m)B(j,l,\alpha)P(z,j|i)\gamma_j(0)^z\right]\right\}.
\end{aligned}
$$

Thus, $\Gamma_t(x_2)$ is composed of vectors $\gamma'(x_2)=\{\gamma'_i(x_2)\}$ of the form,

$$
\begin{aligned}
\gamma'_i(x_2) \;=\; & (c_2-h_2)x_2-\sum_j c_j y_j+L_1(e_i,y_1)+L_2(e_i,y_1,y_2)\\
& +\beta\sum_{j<y_1}\sum_k \overline{P}(k)P(j|i)v(y_1-j,[y_2-k]^+,e_j)\\
& +\beta\sum_z\sum_{x=1}^{y_2}\sum_{j\geq \mathbf{J_1}}\sum_{l=\overline{\mathbf{L}}_{\mathbf{1}}}^{\overline{\mathbf{L}}_{\mathbf{2}}}\overline{P}(y_2-x-l)B(j,l,\alpha)P(z,j|i)\gamma_j(x)^z\\
& +\beta\sum_z\sum_{j\geq \mathbf{J_2}}\sum_{l=\overline{\mathbf{L}}_{\mathbf{3}}}^{\overline{\mathbf{L}}_{\mathbf{4}}}\sum_{m=y_2-l}^{D_2}\overline{P}(m)B(j,l,\alpha)P(z,j|i)\gamma_j(0)^z.
\end{aligned}
$$

We observe that if all of the vectors $\gamma'(x_2)$ are contained in $\Gamma_t(x_2)$, then $|\Gamma_t(x_2)|=(D_1+1)\times\sum_{i=0}^{D_1+D_2}(D_1+D_2+1-\max\{x_2,i\})|\Gamma_{t+1}(i)|^{(D_1+1)}$. Hence the finite representation of $v_t(0,x_2,\xi)$ expands geometrically as $T$ increases..

This discussion suggests that a finite representation of $v_t$ is $(\hat{v}_t,\Gamma_t(0),\cdots,\Gamma_t(D_1+D_2))$, where $v_t(x_1,x_2,e_i)=\hat{v}_t(x_1,x_2,i)$ for all $i$, $x_2$ and $x_1>0$ and $v_t(0,x_2,\xi)=\max\{\xi\gamma:\gamma\in\Gamma_t(x_2)\}$ for all $x_2$ and for all $\xi$. Define the operators $H_1$ and $H_2$ as follows:

$$
H_1(\hat{v},\Gamma(0),\cdots,\Gamma(D_1+D_2))(x_1,x_2,i)=[Hv](x_1,x_2,e_i),
$$

for all $i$, $x_2$ and $x_1>0$, and

$$H_2(x_2)(\hat{v}, \Gamma(0), \cdots, \Gamma(D_1 + D_2))(\xi) = [Hv](0, x_2, \xi),$$

for all $\xi$ and for all $x_2$, where $v(x_1, x_2, e_i) = \hat{v}(x_1, x_2, i)$ for all $i$, $x_2$ and for all $x_1 > 0$ and $v(0, x_2, \xi) = \max\{\xi\gamma : \gamma \in \Gamma(x_2)\}$ for all $\xi$ and for all $x_2$. Then,

$$\hat{v}_t = H_1(\hat{v}_{t+1}, \Gamma_{t+1}(0), \cdots, \Gamma_{t+1}(D_1 + D_2))$$

and

$$\Gamma_t(x_2) = H_2(x_2)(\hat{v}_{t+1}, \Gamma_{t+1}(0), \cdots, \Gamma_{t+1}(D_1 + D_2)).$$

**Sub-Optimal Design**

Although there is a finite representation of $v_t$, the cardinality of the sets of $\gamma$-vectors may still grow prohibitively large as $T$ gets large. We now consider a sub-optimal design that guarantees the cardinality of $\Gamma_t(x_2)$ will never exceed a computable upper bound for all $x_2$.

Recalling that $D_1$ is the maximum demand of product 1, let $a_1(t) = D_1$, independent of $\xi$ and $x_2$, that is selected at time $t$ if $x_1(t) = x_1(t-1) = \cdots x_1(t-K) = 0$ and $x_1(t-K-1) > 0$ for a fixed integer $K \geq 0$ and select $a_2(t)$ using the optimality equation with $a_1(t) = D_1$. Otherwise select both actions that cause the maximum to be obtained in the optimality equation. Thus, once $a_1(t)$ is selected, the inventory level of product 1 at the next decision epoch is guaranteed to be either greater than zero or a special case of zero inventory that allows complete demand observability. That is, note from Equation 9 that if $a_1(t) = D_1$, $x_1(t) = 0$, and $x_1(t+1) = 0$, then $d_1(t+1) = D_1$. Further, demand substitution at the next decision epoch is guaranteed to be 0.

We define:

- $v_t^K(x_1, x_2, e_i)$ as the expected reward to be accrued from $t$ until $T$ under the sub-optimal design policy with parameter $K$ given $\xi(t) = e_i$, $x_1(t) = x_1$ and $x_2(t) = x_2$ where $x_1 > 0$.

- $v_t^k(0, x_2, \xi) \quad \forall \quad k = 1, \cdots, K$ and $\forall \quad x_2 = 0, \cdots, D_1 + D_2$ as the expected reward to be accrued from $t$ until $T$ under the sub-optimal design with parameter $K$ given $\xi(t) = \xi$, $x_2(t) = x_2$, $x_1(t) = x_1(t-1) = \cdots = x(t-K+k) = 0$ and $x(t-K+k-1) > 0$.

- $v_t^0(0, x_2, \xi)$ as the expected reward to be accrued from $t$ until $T$ under the sub-optimal design with parameter $K$ given $\xi(t) = \xi$, $x_2(t) = x_2$, $x_1(t) = x_1(t-1) = \cdots = x_1(t-K) = 0$.

- $\Gamma_t^k(x_2) \quad \forall \quad k = 0, \cdots, K$ and $\forall \quad x_2 = 0, \cdots, D_1 + D_2$ as the set of gamma vectors such that $v_t^k(0, x_2, \xi) = \max_{\gamma \in \Gamma_t^k(x_2)}\{\xi\gamma\}$.

Assume $v_T^K(x_1, x_2, \xi) = \bar{p}_1 x_1 + \bar{p}_2 x_2$ and hence $\Gamma_T^K(x_2) = \{\gamma_T^0(x_2)\}$, where $\gamma_T^0(x_2) = \bar{p}_2 x_2 \mathbf{1}$. Furthermore, let

$$\begin{aligned}
\gamma_{it}^0(x_2) \quad = \quad \max_{y_2 \geq x_2} \Bigg\{ & (c_2 - h_2)x_2 - c_1 y_1^* - c_2 y_2 + L_1(e_i, y_1^*) + L_2(e_i, y_1^*, y_2) \\
& + \beta \sum_{j < y_1^*} \sum_{k} \overline{P}(k) P(j|i) v(y_1^* - j, [y_2 - k]^+, e_j) \qquad (12) \\
& + \beta \sum_{x=1}^{y_2} \sum_{j \geq \mathbf{J_1^*}} \sum_{l=\overline{\mathbf{L}}_\mathbf{1}}^{\overline{\mathbf{L}}_\mathbf{2}^*} \overline{P}(y_2 - x - l) B(j, l, \alpha) P(j|i) \gamma_{j\ t+1}^0(x) \\
& + \beta \sum_{j \geq \mathbf{J_2^*}} \sum_{l=\overline{\mathbf{L}}_\mathbf{3}}^{\overline{\mathbf{L}}_\mathbf{4}^*} \sum_{m=y_2-l}^{D_2} \overline{P}(m) B(j, l, \alpha) P(j|i) \gamma_{j\ t+1}^0(0) \Bigg\},
\end{aligned}$$

where $\mathbf{J_1^*} = y_1^* + [y_2 - x - D_2]^+$, $\mathbf{J_2^*} = y_1^* + [y_2 - D_2]^+$ ,$\overline{\mathbf{L}}_\mathbf{2}^* = \min\{j - y_1^*, y_2 - x\}$, and $\overline{\mathbf{L}}_\mathbf{4}^* = j - y_1^*$.

Let $v_t^K = v_t,\ t = T - K, \cdots, T$.

**Algorithm 4**

For $t < T - K$, assume the array

$$(v_{t+k}^K,\ k = 1, \cdots K,\ \Gamma_{t+1}^K(0), \cdots, \Gamma_{t+1}^K(D_1 + D_2), \gamma_{t+K}^0(0), \cdots, \gamma_{t+K}^0(D_1 + D_2))$$

is given, where $v_{t+k}^K = \{v_{t+k}^K(x_1, x_2, e_i) : x_1 > 0\},\ k = 1, \cdots, K$. We remark that this array fully determines $v_{t+1}^K(x_1, x_2, \xi)$ for all $x_1$, $x_2$ and $\xi$. We determine

$$(v_{t+k}^K,\ k = 0, \cdots K-1,\ \ \Gamma_t^K(0), \cdots, \Gamma_t^K(D_1+D_2), \gamma_{t+K-1}^0(0), \cdots, \gamma_{t+K-1}^0(D_1+D_2))$$

as follows:

(i) $\gamma_{t+K-1}^0(x)$ is determined from $v_{t+K}^K$ and $\gamma_{t+K}^0(0), \cdots, \gamma_{t+K}^0(D_1+D_2)$.

(ii) $v_t^K = H_1(v_{t+1}^K, \Gamma_{t+1}^K(0), \cdots, \Gamma_{t+1}^K(D_1+D_2))$.

(iii) $\Gamma_{t+k}^{K-k}(x) = H_2(x)(v_{t+k+1}^K, \Gamma_{t+k+1}^{K-k-1}(0), \cdots, \Gamma_{t+k+1}^{K-k-1}(D_1+D_2))$ for $k = 0, \cdots, K-1$.

We note that the cardinality of the array $(v_{t+k}^K,\ k = 0, \cdots K-1,\ \Gamma_t^K(0), \cdots, \Gamma_t^K(D_1+D_2), \gamma_{t+K-1}^0(0), \cdots, \gamma_{t+K-1}^0(D_1+D_2))$ is $K \times D_1 \times (D_1+1) \times (D_1+D_2+1) + \sum_{i=0}^{D_1+D_2} |\Gamma_t^K(i)| + (D_1+D_2+1)$, where $|\Gamma_t^0(x)| = 1$ for all $x$, and $|\Gamma_t^k(x)| \leq (D_1+1) \times \sum_{i=0}^{D_1+D_2}(D_1+D_2+1-\max\{x,i\})|\Gamma_{t+1}^{k-1}(i)|^{(D_1+1)}$. Further, we note that transition from $(v_{t+1}^K, \cdots, \gamma_{t+K}(D_1+D_2))$ to $(v_t^K, \cdots, \gamma_{t+K-1}(D_1+D_2))$ requires application of the $H_2$ operator $K \times (D_1+D_2+1)$ times.

We explain the use of the $\gamma_t^0(x)$ vector as follows. For simplicity, let $K = 0$; that is, assume we order $y_1^*$ items of product 1 whenever its inventory goes to zero, irrespective of $\xi$ and $x_2$ (in reality not a particulary clever sub-optimal design). Let $v_t^0$ be the resulting expected value to be accrued from $t$ till $T$. Then,

$$\begin{aligned} v_t^0(0, x_2, \xi) \quad = \quad \max_{y_2 \geq x_2} \Bigg\{ & (c_2 - h_2)x_2 - c_1 y_1^* - c_2 y_2 + L_1(e_i, y_1^*) + L_2(e_i, y_1^*, y_2) + \\ & \beta \sum_{j < y_1^*} \sum_i \xi_i \sum_k \overline{P}(k) P(j|i) v(y_1^* - j, [y_2 - k]^+, e_j) + \\ & \beta \sum_z \sum_{x=1}^{y_2} \sigma^1(z, \xi, y_1^*, y_2, x) v_{t+1}^0(0, x, \lambda^1(z, \xi, y_1^*, y_2, x)) + \\ & \beta \sum_z \sigma^2(z, \xi, y_1^*, y_2) v_{t+1}^0(0, 0, \lambda^2(z, \xi, y_1^*, y_2)) \Bigg\}. \end{aligned}$$

We recall that $v_T^0(0, x_2, \xi) = \bar{p}_2 x_2$; hence, $\Gamma_T^0(x_2) = \{\bar{p}_2 x_2 \mathbf{1}\}$. Assume $\Gamma_{t+1}^0(x_2)$ is also a singleton for all $x_2$; *i.e.*, $\Gamma_{t+1}^0(x_2) = \{\gamma_{t+1}^0(x_2)\}$ for all $x_2$. Then,

$$\sum_z \sum_{x=1}^{y_2} \sigma^1(z,\xi,y_1^*,y_2,x) v_{t+1}^0(0,x,\lambda^1(z,\xi,y_1^*,y_2,x)) =$$

$$\sum_z \sum_{x=1}^{y_2} \sigma^1(z,\xi,y_1^*,y_2,x)\lambda^1(z,\xi,y_1^*,y_2,x)\gamma_{t+1}^0(x)$$

$$= \sum_z \sum_{x=1}^{y_2} \sum_{j \geq \mathbf{J_1^*}} \sum_{l=\overline{\mathbf{L}}_\mathbf{1}}^{\overline{\mathbf{L}}_\mathbf{2}^*} \overline{P}(y_2 - x - l) B(j,l,\alpha) \sum_i \xi_i P(z,j|i) \gamma_{j\ t+1}^0(x)$$

$$= \sum_{x=1}^{y_2} \sum_{j \geq \mathbf{J_1^*}} \sum_{l=\overline{\mathbf{L}}_\mathbf{1}}^{\overline{\mathbf{L}}_\mathbf{2}^*} \overline{P}(y_2 - x - l) B(j,l,\alpha) \sum_i \xi_i P(j|i) \gamma_{j\ t+1}^0(x),$$

where the last equality is due to the fact that $\sum_z P(z,j|i) = P(j|i)$ and that $\gamma_{t+1}^0(x)$ is independent of $z$. Similarly,

$$\sum_z \sigma^2(z,\xi,y_1^*,y_2,) v_{t+1}^0(0,0,\lambda^2(z,\xi,y_1^*,y_2)) =$$

$$\sum_z \sigma^2(z,\xi,y_1^*,y_2,)\lambda^2(z,\xi,y_1^*,y_2)\gamma_{t+1}^0(0)$$

$$= \sum_z \sum_{j \geq \mathbf{J_2^*}} \sum_{l=\overline{\mathbf{L}}_\mathbf{3}}^{\overline{\mathbf{L}}_\mathbf{4}^*} \sum_{m=y_2-l}^{D_2} \overline{P}(m) B(j,l,\alpha) \sum_i \xi_i P(z,j|i) \gamma_{j\ t+1}^0(0)$$

$$= \sum_{j \geq \mathbf{J_2^*}} \sum_{l=\overline{\mathbf{L}}_\mathbf{3}}^{\overline{\mathbf{L}}_\mathbf{4}^*} \sum_{m=y_2-l}^{D_2} \overline{P}(m) B(j,l,\alpha) \sum_i \xi_i P(j|i) \gamma_{j\ t+1}^0(0).$$

Thus, if $\Gamma_{t+1}^0(x)$ is a singleton for all $x$ and the action $a_1$ taken is $\xi$-invariant, then $\Gamma_t^0(x)$ is also an (easily computed) singleton for all $x$. It seems reasonable that $v_t^{K+1}$ would be at least as good an approximation as $v_t^K$, as stated in Proposition 4 (proof can be found in Appendix **??**).

**Proposition 4** *For all t,* $v_t^K \leq v_t^{K+1} \leq v_t$.

We now present an alternative approach for determining $v_t^K$.

**Algorithm 5**

For $t < T - K$ assume the array $(v_{t+1}^K, \Gamma_{t+1}^k(x), k = 0, \cdots, K, x = 0, \cdots, (D_1 + D_2))$ is given, where $v_{t+1}^K = \{v_{t+1}^K(x_1, x_2, e_i) : x_1 > 0\}$. We remark that this array fully determines $v_{t+1}^K(x_1, x_2, \xi)$ for all $x_1$, $x_2$ and $\xi$. We determine $(v_t^K, \Gamma_t^k(x), k = 0, \cdots, K, x = 0, \cdots, (D_1 + D_2))$ as follows:

(i) $\gamma_t^0(x)$ is determined from $v_{t+1}^K$ and $\gamma_{t+1}^0(0), \cdots, \gamma_{t+1}^0(D_1 + D_2)$.

(ii) $v_t^K = H_1(v_{t+1}^K, \Gamma_{t+1}^K(0), \cdots, \Gamma_{t+1}^K(D_1 + D_2))$.

(iii) $\Gamma_t^k(x) = H_2(x)(v_{t+1}^K, \Gamma_{t+1}^{k-1}(0), \cdots, \Gamma_{t+1}^{k-1}(D_1 + D_2))$ for $k = 1, \cdots, K$.

We note that the cardinality of the array $(v_t^K, \Gamma_t^k(x), k = 0, \cdots, K, x = 0, \cdots D_1 + D_2)$ is $D_1 \times (D_1 + D_2 + 1) \times (D_1 + 1) + \sum_{k=0}^{K} \sum_{x=0}^{D_1+D_2} |\Gamma_t^k(x)|$, where $|\Gamma_t^0| = 1$ and $|\Gamma_t^k(x)| \leq (D_1 + 1) \times \sum_{i=0}^{D_1+D_2} (D_1 + D_2 + 1 - \max\{x, i\}) |\Gamma_{t+1}^{k-1}(i)|^{(D_1+1)}$. Further, we note that transition from $(v_{t+1}^K, \Gamma_{t+1}^k(x), k = 0, \cdots, K, x = 0, \cdots, D_1 + D_2)$ to $(v_t^K, \Gamma_t^k(x), k = 0, \cdots, K, x = 0, \cdots, D_1 + D_2)$ requires application of the $H_2$ operator $K \times (D_1 + D_2 + 1)$ times.

We remark that on the basis of operations count, Algorithm 4 would be preferred to, Algorithm 5. However, as we will now show, Algorithm 5 suggests an algorithm, Algorithm 6 presented below, that is based on a non-standard sufficient statistic offering a significantly simpler approach for software development.

### 4.5.2 Approach 2

The first approach for constructing an optimal policy for the *IRPDS* was based on the fact that for finite $T$, $v_t(0, x_2, \xi)$ has a finite representation, $\{\Gamma_t(0), \cdots, \Gamma_t(D_1 + D_2)\}$, although $\xi$ is a member of an uncountably infinite set. The second approach for constructing an optimal policy is based on the fact that $|\mathcal{H}(t)|$ is finite for finite $t$. We also make use of the fact that there exists a set $\mathcal{H}'(t) \subseteq \mathcal{H}(t)$ that can also serve as a sufficient statistic for the *IRPDS*, where $\mathcal{H}'(t) = \{z_1(t), \cdots, z_1(t - \tau + 1), a_1(t - 1), \cdots, a_1(t - \tau), a_2(t - 1), \cdots, a_2(t - \tau), x_2(t), \cdots, x_2(t - \tau + 1), x_1(t - \tau), d_1(t - \tau)\}$, and where $\tau$ is such that

$x_1(t) = x_1(t-1) = \cdots = x_1(t-\tau+1) = 0$ and $x_1(t-\tau) > 0$. Proof of the following result, which justifies the claim that $\mathcal{H}'$ is a sufficient statistic, is due to the fact that $x_1(t) > 0$ implies $d_1(t)$ is completely observed.

**Proposition 5** *For all t,* $P(d_1(t) = i|\mathcal{H}'(t)) = P(d_1(t) = i|\mathcal{H}(t))$.

Let $\mathcal{H}_0 = \{(x_1, x_2, e_i) : x_1 > 0, i \in \{0, \cdots, D_1\}, x_2 \in \{0, \cdots, D_1 + D_2\}\}$, $\mathcal{H}_1(0) = \{\lambda^2(z, \xi, y_1, y_2) : (x_1, x_2, \xi) \in \mathcal{H}_0, y_1 \in \{x_1, \cdots, D_1\}, y_2 \in \{x_2, \cdots, D_1 + D_2\}, z \in \{0, 1, \cdots, D_1\}\}$ and $\mathcal{H}_1(x) = \{\lambda^1(z, \xi, y_1, y_2, x) : (x_1, x_2, \xi) \in \mathcal{H}_0, y_1 \in \{x_1, \cdots, D_1\}, y_2 \in \{x_2, \cdots, D_1 + D_2\}, z \in \{0, 1, \cdots, D_1\}\}$ for all $x \in \{1, \cdots, D_1 + D_2\}$. For $k \geq 1$, let $\mathcal{H}_{k+1}(x) = \{\lambda^1(z, \xi, y_1, y_2, x) : \xi \in \bigcup_i \mathcal{H}_k(i),\ z, y_1 \in \{0, \cdots, D_1\},\ y_2 \in \{i, \cdots, D_1 + D_2\}\}$ for all $x \in \{1, \cdots, D_1 + D_2\}$. Also let, $\mathcal{H}_{k+1}(0) = \{\lambda^2(z, \xi, y_1, y_2) : \xi \in \bigcup_i \mathcal{H}_k(i),\ z, y_1 \in \{0, \cdots, D_1\}; y_2 \in \{i, \cdots, D_1 + D_2\}\}$. Further, let $\mathcal{H}_k = \bigcup_x \mathcal{H}_k(x)$. We remark that $\mathcal{H}_k$ is equivalent to $\mathcal{H}'(t)$, given $\tau = k$. As a slight abuse of notation, let $H_1(v, \tilde{v}(0, 0, \cdot), \cdots, \tilde{v}(0, D_1 + D_2, \cdot)) = H_1(v, \Gamma(0), \cdots, \Gamma(D_1 + D_2))$ and $H_2(x)(v, \tilde{v}(0, 0, \cdot), \cdots, \tilde{v}(0, D_1 + D_2, \cdot)) = H_2(x)(v, \Gamma(0), \cdots, \Gamma(D_1 + D_2))$ if $\tilde{v}(0, x, \xi) = \max\{\xi\gamma : \gamma \in \Gamma(x)\}$. We now present an algorithm for determining $v_t^K(x_1, x_2, \xi)$ for all $(x_1, x_2, \xi) \in \mathcal{H}_0$ and $v_t^{K-k+1}(0, x, \xi)$ for all $k = 1, \cdots, K+1$, for all $x = 0, \cdots, D_1 + D_2$ and for all $(0, x, \xi)$ such that $\xi \in \mathcal{H}_k(x)$.

**Algorithm 6**

For $t < T - K$, assume $v_{t+1}^K(x_1, x_2, \xi)$ is given, for all $(x_1, x_2, \xi) \in \mathcal{H}_0$ and $v_{t+1}^{K-k+1}(0, x_2, \xi)$ is given for all $k = 1, \cdots, K+1$, for all $x_2 = 0, \cdots, D_1 + D_2$ and for all $(0, x_2, \xi)$ such that $\xi \in \mathcal{H}_k(x_2)$, where $v_{t+1}^0(0, x_2, \xi) = \xi\gamma_{t+1}^0(x_2)$ for all $\xi \in \mathcal{H}_{K+1}(x_2)$ and for all $x_2 = 0, \cdots, D_1 + D_2$, also assume $\gamma_{t+1}^0(x_2)$ is given for all $x_2 = 0, \cdots, D_1 + D_2$. Let $\overline{v}_{t+1} = \{v_{t+1}^K(x_1, x_2, \xi) : (x_1, x_2, \xi) \in \mathcal{H}_0\}$. We determine $v_t^K(x_1, x_2, \xi)$ for all $(x_1, x_2, \xi) \in \mathcal{H}_0$ and $v_t^{K-k+1}(0, x_2, \xi)$ for all $k = 1, \cdots, K+1$, for all $x_2 = 0, \cdots, D_1 + D_2$ and for all $(0, x_2, \xi)$ such that $\xi \in \mathcal{H}_k(x_2)$, where $v_t^0(0, x_2, \xi) = \xi\gamma_t^0(x_2)$ for $\xi \in \mathcal{H}_{K+1}(x_2)$ as follows:

(i) $v_t^K(x_1, x_2, \xi) = H_1(\overline{v}_{t+1}, v_{t+1}^K(0, 0, \cdot), \cdots, v_{t+1}^K(0, D_1 + D_2, \cdot))(x_1, x_2, \xi)$
for all $(x_1, x_2, \xi) \in \mathcal{H}_0$ where $v_{t+1}(0, x, \cdot) = \{v_{t+1}^K(0, x, \xi) : \xi \in \mathcal{H}_1(x)\}$.

(ii) $v_t^{K-k+1}(0, x_2, \xi) = H_2(\overline{v}_{t+1}, v_{t+1}^{K-k}(0, 0, \cdot), \cdots, v_{t+1}^{K-k}(0, D_1 + D_2, \cdot))(0, x_2, \xi)$ for all $\xi \in \mathcal{H}_k(x_2)$, where $v_{t+1}^{K-k}(0, x, \cdot) = \{v_{t+1}^{K-k}(0, x, \xi) : \xi \in \mathcal{H}_{k+1}(x)\}$ for $k = 1, \cdots, K$ and $x_2 = 0, \cdots, D_1 + D_2$.

(iii) $v_t^0(0, x_2, \xi) = \xi\gamma_t^0(x_2)$, where $\gamma_{it}^0(x_2)$ is obtain using equation (12).

We remark that Algorithm 5 and 6 are nearly identical, differing only as follows:

(i) The algorithms use different representations of the $v_t^k(0, x_2, \cdot)$ functions, where the representation in Algorithm 6 is significantly simpler for software implementation than is the representation in Algorithm 5

(ii) Algorithm 6 holds for all $(x_1, x_2, \xi) \in \mathcal{H}_0$, for all $x_2 = 0, \cdots, D_1 + D_2$ and for all $(0, x_2, \xi)$ such that $\xi \in \mathcal{H}_1(x_2) \bigcup \cdots \bigcup \mathcal{H}_K(x_2)$ whereas Algorithm 5 holds for all $(x_1, x_2, \xi) \in \mathcal{H}_0$, for all $x_2 = 0, \cdots, D_1 + D_2$ and for all $(0, x_2, \xi)$ such that $\xi \in \Xi$.

### *4.6 A Method for Bounding the Value of Demand Observability*

We now present a procedure that uses the previously described model and solution approaches to bound the maximum value of improved demand observability for the *IRPDS*. To do so, we consider two extreme cases that we call *completely-observed* and *sales-only-observed*. In the completely-observed case, we assume that the observation process provides a perfect observation of demand of product 1 in the prior period, even when $x_1(t) = 0$. Let $V_o^*(x_1, x_2, \xi)$ denote the value of maximum expected profit over some fixed planning horizon for this case, given $x_1(0) = x_1$, $x_2(0) = x_2$ and $\xi^1(0) = \xi$. At the other extreme, the sales-only-observed case assumes that the observation process provides no additional information about demand of product 1. Therefore, order decisions are made using only information obtained from sales data. Since this case corresponds to the situation in which improving demand observability provides no benefit, it will be useful for developing a lower bound. Let $V_s^*(x_1, x_2, \xi)$ denote the value of maximum expected profit in this case, again given $x_1(0) = x_1$, $x_2(0) = x_2$ and $\xi^1(0) = \xi$. Results in White and Harrington [55] guarantee that $V_o^*(x_1, x_2, \xi) \geq V_s^*(x_1, x_2, \xi)$, for all $(x_1, x_2, \xi)$.

Clearly, the gap between these two values, $V_o^*(x_1, x_2, \xi) - V_s^*(x_1, x_2, \xi)$ corresponds to the *maximum* added expected benefit that can result through the application of techniques that aim at improving demand observability.

As described earlier, the complicating feature of the operator $H$ is due to the partial observability of the demand of product 1. Therefore, determining an optimal policy and the resultant maximum expected profit for the completely-observed case does not require the use of suboptimal techniques for problem settings of reasonable size. On the other hand, it will usually be computationally prohibitive to determine an optimal policy for the sales-only-observed case, so instead we use the suboptimal solution approaches developed in Section 4.5. Let $V_s^{LB(K)}(x_1, x_2, \xi)$ denote a lower bound for $V_s^*(x_1, x_2, \xi)$, obtained by applying the suboptimal design with parameter $K$. Thus,

$$V_o^*(x_1, x_2, \xi) - V_s^{LB(K)}(x_1, x_2, \xi) \tag{13}$$

corresponds to an upper bound on the maximum added value that can be obtained from improving demand observability. Of course, larger values of $K$ lead to tighter lower bounds $V_s^{LB(K)}(x_1, x_2, \xi)$ which in turn lead to tighter upper bounds on the maximum value due to improved demand observability.

### 4.6.1 Completely-Observed Case

To analyze the completely-observed case of the *IRPDS*, we simply assume that $P(z|j, i) = 1$ if and only if $z = j$. Thus, $z_1(t)$ is a perfect observation of $d_1(t)$ (independent of the value of $x_1(t)$ and $x_2(t)$). Hence, the only general state of interest is now $(x_1, x_2, e_i)$, $x_1 \geq 0$ and $x_2 \geq 0$. Let

$$
\begin{aligned}
h'(i, y_1, y_2, v) &= -\sum_j c_j y_j + L_1(i, y_1) + L_2(i, y_1, y_2) \\
&\quad +\beta \sum_{j<y_1} \sum_k \overline{P}(k) P(j|i) v(y_1 - j, [y_2 - k]^+, j) \\
&\quad +\beta \sum_{x=1}^{y_2} \sum_{j \geq \mathbf{J_1}} \sum_{l=\overline{\mathbf{L}}_\mathbf{1}}^{\overline{\mathbf{L}}_\mathbf{2}} \overline{P}(y_2 - x - l) B(j, l, \alpha) P(j|i) v(0, x, j) \\
&\quad +\beta \sum_{j \geq \mathbf{J_2}} \sum_{l=\overline{\mathbf{L}}_\mathbf{3}}^{\overline{\mathbf{L}}_\mathbf{4}} \sum_{m \geq [y_2 - l]^+} \overline{P}(m) B(j, l, \alpha) P(j|i) v(0, 0, j)
\end{aligned}
$$

and

$$
[H'v'](x_1, x_2, i) = \sum_j (c_j - h_j) x_j + \max_{\substack{y_1 \geq x_1 \\ y_2 \geq x_2}} h'(i, y_1, y_2, v).
$$

Then,

$$
v'_t = H' v'_{t+1} \tag{14}
$$

is the optimality equation for the completely-observed case, where the boundary condition is $v'_T(x_1, x_2, i) = \bar{p}_1 x_1 + \bar{p}_2 x_2$ for all $i$.

### 4.6.2 Sales-Only-Observed Case

To analyze the case where demand is only observable through sales data, we assume that $P(z|j, i)$ is independent of $i$ and $j$. Thus, $z_1(t)$ contains no information about the value of $d_1(t)$. The general optimality equation now becomes $v''_t = H'' v''_{t+1}$, $v''_T(x_1, x_2, \xi) = \bar{p}_1 x_1 + \bar{p}_2 x_2$ for all $\xi$, where:

$$
[H''v''](x_1, x_2, \xi) = \sum_j (c_j - h_j) x_j + \max_{\substack{y_1 \geq x_1 \\ y_2 \geq x_2}} h''(\xi, y_1, y_2, v),
$$

$$
\begin{aligned}
h''(\xi, y_1, y_2, v) &= -\sum_i c_i y_i + L_1(\xi, y_1) + L_2(\xi, y_1, y_2) + \\
&\beta \sum_j < y_1 \sum_i \xi_i \sum_k \overline{P}(k) P(j|i) v(y_1 - j, [y_2 - k]^+, j) + \\
&\beta \sum_{x=1}^{y_2} \overline{\sigma}^1(\xi, y_1, y_2, x) v(0, x, \overline{\lambda}^1(\xi, y_1, y_2, x)) + \\
&\beta \overline{\sigma}^2(\xi, y_1, y_2) v(0, 0, \overline{\lambda}^2(\xi, y_1, y_2)),
\end{aligned}
$$

$$
\overline{\sigma}^1(\xi, y_1, y_2, x) = \sum_{j \geq \mathbf{J_1}} \sum_{l=\overline{\mathbf{L}}_\mathbf{1}}^{\overline{\mathbf{L}}_\mathbf{2}} \overline{P}(y_2 - x - l) B(j, l, \alpha) \sum_i \xi_i P(j|i),
$$

$$
\overline{\lambda}_j^1(\xi, y_1, y_2, x) = \frac{\sum_{l=\overline{\mathbf{L}}_\mathbf{1}}^{\overline{\mathbf{L}}_\mathbf{2}} \overline{P}(y_2 - x - l) B(j, l, \alpha) \sum_i \xi_i P(j|i)}{\overline{\sigma}^1(\xi, y_1, y_2, x)} \quad \text{if} j \geq \mathbf{J_1},
$$

and $\overline{\lambda}_j^1(\xi, y_1, y_2, x) = 0$ otherwise.

$$
\overline{\sigma}^2(\xi, y_1, y_2) = \sum_{j \geq \mathbf{J_2}} \sum_{l=\overline{\mathbf{L}}_\mathbf{3}}^{\overline{\mathbf{L}}_\mathbf{4}} \sum_{m \geq [y_2 - l]^+} \overline{P}(m) B(j, l, \alpha) \sum_i \xi_i, P(j|i),
$$

$$
\overline{\lambda}_j^2(\xi, y_1, y_2) = \frac{\sum_{l=\overline{\mathbf{L}}_\mathbf{3}}^{\overline{\mathbf{L}}_\mathbf{4}} \sum_{m \geq [y_2 - l]^+} \overline{P}(m) B(j, l, \alpha) \sum_i \xi_i P(j|i)}{\overline{\sigma}^2(\xi, y_1, y_2)} \quad \text{if} j \geq \mathbf{J_2},
$$

and $\overline{\lambda}_j^2(\xi, y_1, y_2) = 0$ otherwise.

### 4.6.3 Computing a Bound

For the completely-observed case, we use expression (14) to determine $V_o^*(x_1, x_2, \xi) = v_0'(x_1, x_2, \xi)$ for all potential initial states. It is important to note that these states include all $(x_1, x_2, \xi) \in \mathcal{H}_0$, as well as states $(0, x_2, e_i)$ for $i = 0, \cdots, D_1$ and for $x_2 = 0, \cdots, D_1 + D_2$. For the sales-only-observed case, we use suboptimal Algorithm 6 to determine $V_s^{LB(K)}(x_1, x_2, \xi) \leq v_0''(x_1, x_2, \xi)$ because it is easier to implement in software relative to Algorithms 4 and 5. In this case, the potential initial states again include all $(x_1, x_2, \xi) \in \mathcal{H}_0$, but we restrict the zero inventory states to those that might be visited by Algorithm 6; *i.e.*, all states $(0, x_2, \xi)$ such that $\xi \in \mathcal{H}_1'' \bigcup \cdots \bigcup \mathcal{H}_{K+1}''$, where $\mathcal{H}_k'' = \bigcup_i \mathcal{H}''(i)_k$, $\mathcal{H}_1''(0) = \{\overline{\lambda}^2(\xi, y_1, y_2) : (x_1, x_2, \xi) \in \mathcal{H}_0, y_1 \in \{x_1, \cdots, D_1\}, y_2 \in \{x_2, \cdots, D_1 + D_2\}\}$, $\mathcal{H}_{k+1}''(0) =$

$\{\overline{\lambda}^2(\xi, y_1, y_2) : \xi \in \bigcup_i \mathcal{H}_k(i), y_1 \in \{0, \cdots, D_1\}, y_2 \in \{i, \cdots, D_1 + D_2\}\}$, and for $x > 0$, $\mathcal{H}''_1(x) = \{\overline{\lambda}^1(\xi, y_1, y_2, x) : (x_1, x_2, \xi) \in \mathcal{H}_0, y_1 \in \{x_1, \cdots, D_1\}, y_2 \in \{x_2, \cdots, D_1 + D_2\}\}$ and $\mathcal{H}''_{k+1}(x) = \{\overline{\lambda}^1(\xi, y_1, y_2, x) : \xi \in \bigcup_i \mathcal{H}_k(i), y_1 \in \{0, \cdots, D_1\}, y_2 \in \{i, \cdots, D_1 + D_2\}\}$.

Given an initial state $(x_1, x_2, \xi) \in \mathcal{H}_0$ that is shared by both cases, it is possible to calculate a bound on the expected added benefit of improving demand observability using $V_o^*(x_1, x_2, \xi) - V_s^{LB(K)}(x_1, x_2, \xi)$. For initial states $(0, x_2, \xi)$, a similar bound can be computed by blending the value function using the prior distribution. To do so, let $V_o^*(0, x_2, \xi) = \sum_i \xi_i v'_0(0, x_2, e_i)$. Then $V_o^*(0, x_2, \xi) - V_s^{LB(K)}(0, x_2, \xi)$ again represents a bound.

Since it may be useful to determine a measure for the potential value of observability that is independent of the initial state, we note that a reasonable approach may be to compare a weighted sum of the state-wise maximum percentage potential gains due to improved demand observability, where the weights given to each state correspond to its likelihood of visitation. Given scalar weights $w^{x_2,\xi} \geq 0$ for all $\xi \in \mathcal{H}''_1 \bigcup \cdots \bigcup \mathcal{H}''_{K+1}$ and $w^{x_1,x_2,\xi} \geq 0$ for all $(x_1, x_2, \xi) \in \mathcal{H}_0$ such that $\sum w^{x_2,\xi} + \sum w^{x_1,x_2,\xi} = 1$, such a weighted statistic is given by the following expression:

$$
\begin{aligned}
G^{UB(K)}(w) \;=\; & \sum_{x_2} \sum_{\xi \in \mathcal{H}''_1(x_2) \bigcup \cdots \bigcup \mathcal{H}''_{K+1}(x_2)} w^{x_2,\xi} \frac{\left(V_o^*(0, x_2, \xi) - V_s^{LB(K)}(0, x_2, \xi)\right)}{V_s^{LB(K)}(0, x_2, \xi)} \\
& + \sum_{(x_1,x_2,\xi) \in \mathcal{H}_0} w^{x_1,x_2,\xi} \frac{\left(V_o^*(x_1, x_2, \xi) - V_s^{LB(K)}(x_1, x_2, \xi)\right)}{V_s^{LB(K)}(x_1, x_2, \xi)}.
\end{aligned} \tag{15}
$$

### *4.7 Computational Analysis*

Now we apply our methodology for bounding the value of demand observability to a set of problem scenarios. The main objective of the analysis presented in this section is to develop a better understanding of the impact that product substitution has on the value of improved demand observability. Our experiments also examine the effect of two problem characteristics on the performance of the suboptimal policy for the sales-only-observed case: (1) similarity of the products, measured by the ratio of their profits, and (2) the value of

the parameter $K$ in the suboptimal design.

Three sets of scenarios were generated in order to address the objectives of this study. In all scenarios, we assume that product type 2 is preferred over product type 1, but is more expensive. Thus, customers who originally would like to purchase product type 1 do so because of budget concerns, and may be willing to substitute preferred product type 2 at higher cost.

To mitigate the impact of initial conditions on the experiments, we employ a long maximum planning horizon of $T = 1000$ periods for each scenario, and the same stopping criteria described in Section 3.7 of Chapter 3. The maximum number of iterations required to solve the scenarios presented in this section was 343 ($< T$). Therefore, it is also true that the stationary policies (and their resultant state-wise expected profits) obtained in the final iteration are good approximations of the results that would be found for the infinite horizon problem formulations.

In all scenarios, the Markovian demand process for product type 1 is dependent on two parameters, $\zeta$ and $r$. Parameter $\zeta$ is the probability that the demand level in period $t+1$ is the same as the demand in period $t$, and the maximum amount by which the demand level may change from one period to the next is a function of parameter $r$. See Section 3.7 of Chapter 3 for a complete description of these parameters. The demand distribution for product type 2 is assumed to be discrete uniform on the interval $[0, D_2]$. The end-of-horizon inventory salvage values $\bar{p}_1$ and $\bar{p}_2$ where set to zero.

For each problem scenario, we calculate state-wise expected profit for the completely-observed case $V_o^*(x_1, x_2, e_i)$ for all $x_1, x_2 \geq 0$ and $i \in \{0, \cdots, D_1\}$ using equation 14. Next we use suboptimal Algorithm 6 with parameter $K$ to determine $V_s^L B(K)(x_1, x_2, \xi)$ for all $(x_1, x_2, \xi) \in \mathcal{H}_0$ and all $(0, x, \xi)$ such that $\xi \in \mathcal{H}_1'' \bigcup \cdots \bigcup \mathcal{H}_{K+1}''$. For all $(x_1, x_2, \xi) \in \mathcal{H}_0$ and $(0, x, \xi)$ such that $\xi \in \mathcal{H}_1'' \bigcup \cdots \bigcup \mathcal{H}_{K+1}''$ let:

- $a_t^{LB}(x_1, x_2, \xi)$ be the decision rule at time $t$ found using the sub-optimal design in the sales-only-observed case, given $x_1(t) = x_1$, $x_2(t) = x_2$ and $\xi(t) = \xi$;
- $P^t_{(x_1,x_2,\xi),(\overline{x}_1,\overline{x}_2,\overline{\xi})}(a_t^{LB}(x_1, x_2, \xi)) = P(x_1(t+1) = \overline{x}_1, x_2(t+1) = \overline{x}_2, \xi(t+1) = \overline{\xi} | x_1(t) =$

$x_1, x_2(t) = x_2, \xi(t) = \xi, a(t) = a_t^{LB}(x_1, x_2, \xi))$ be the transition probability from state $(x_1, x_2, \xi)$ to state $(\overline{x}_1, \overline{x}_2, \overline{\xi})$ at time $t$ given the sub-optimal decision rule;

- and $P(t) = \{P^t_{(x_1,x_2,\xi),(\overline{x}_1,\overline{x}_2,\overline{\xi})}(a_t^{LB})\}$ be the corresponding transition probability matrix for time $t$ given the sub-optimal decision rule.

As mentioned earlier, due to the length of the planning horizon and the utilized stopping criterion, a good approximation of the steady-state probabilities for all $(x_1, x_2, \xi) \in \mathcal{H}_0$ (observe that $(x_1, x_2, \xi) \in \mathcal{H}_0$ implies $\xi \in \{e_0, \cdots, e_{D_1}\}$) and $(0, x, \xi)$ such that $\xi \in \mathcal{H}''_1 \bigcup \cdots \bigcup \mathcal{H}''_{K+1}$ is obtained by calculating the steady-state probabilities of the Markov chain associated with the stochastic matrix $P(0)$. We use these steady-state probabilities as the weights $w^{x_1,x_2,\xi}$ and $w^{x_2,\xi}$ to compute the maximum percentage value statistic given by (15).

The objective of the first experiment is to determine the impact that product similarity has on the performance of our suboptimal policy. Product similarity is measured as the ratio between the per unit profit of product 1 and the per unit profit of product 2, $\frac{p_1-c_1}{p_2-c_2}$, which we refer to as Product Similarity Ratio ($PSR$). The per unit ordering cost, selling price and holding cost of product type 2 are set to be $c_2 = 160$, $p_2 = 200$ and $h_2 = 5$ respectively. Maximum demands for products type 1 and type 2 are set to be $D_1 = 4$ and $D_2 = 2$ respectively. The discount factor and the parameters of the Markovian demand process of product type 1 are set to be $\beta = 0.95$, $r = 2$ and $\zeta = 0.85$. Different levels of substitution are considered, and the probability of substitution $\alpha$ is selected from the set $\{0, 0.2, 0.25, 0.3, 0.5, 0.6, 0.75, 0.8, 0.9, 1\}$. The parameter value for the suboptimal policy chosen was $K = 2$. The per unit ordering cost, selling price and holding cost of product type 1 is varied (see Table 2) in order to account for different levels of product similarity.

Our numerical experiments indicate that product similarity has a significant effect on the quality of the bound of the value of improved demand observability. As depicted in Figure 16, for a specific level of substitutability ($\alpha$) the bound on the value of increased demand observability increases as $PSR$ decreases. Further, for any given level of $PSR$, there exists a breakpoint value $\alpha$ such that the maximum percentage improvement due to

increased demand observability increases significantly to the right of the breakpoint. For instance, when $PSR = 0.25$, this breakpoint occurs at $\alpha = 0.3$. This suggests that for every level of $PSR$, the suboptimal policy performs well only up to some critical value of $\alpha$. For a given level of similarity ($PSR$), the optimal ordering amount of product type 1 should decrease as substitutability ($\alpha$) increases. Since the suboptimal policy will choose to order a large quantity after a few successive periods of product 1 stockout, it should be clear that the quality of the solution given by the suboptimal policy degrades as $\alpha$ increases. In this experiment, we observe that the proposed suboptimal policy performs reasonably well in systems in which demand substitutability is smaller than $PSR$. We remark that most typical real-world supply chain systems should have such characteristics. It is very unlikely to find a high percentage of customers willing to substitute their initial choice with a product with a significantly higher price.

The objective of the second experiment is to understand the effect that parameter $K$ has on the performance of the suboptimal policy. We focus our analysis in a system with level of substitutability $\alpha = 0.8$ and two levels of $PSR$, 0.75 and 0.25. The maximum demand of products type 1 and type 2 are set to be $D_1 = 2$ and $D_2 = 2$ respectively. The discount factor and the parameters of the Markovian demand process of product type 1 are set to be $\beta = 0.95$, $r = 1$ and $\zeta = 0.85$. The value of the parameter of the suboptimal policy $K$ is varied between scenarios, and is selected from the set $\{1, 2, 3, 4\}$. The per unit ordering cost and the holding cost of products type 1 and type 2 are set to be $c_1 = 120$, $h_1 = 3$, $c_2 = 160$, and $h_2 = 5$. The selling price of product type 2 is set to be $p_2 = 200$. The selling price of product type 1 is selected form the set $\{150, 130\}$.

From the numerical experiments we observe that when $PSR = 0.75$ the value of $K$ does

**Table 2:** Parameters for Product Type 1

| $c_1$ | $p_1$ | $h_1$ | $PSR$ |
|---|---|---|---|
| 120 | 120 | 0.000001 | 0 |
| 120 | 130 | 1 | 0.25 |
| 120 | 140 | 2 | 0.5 |
| 120 | 150 | 3 | 0.75 |
| 120 | 160 | 4 | 1 |

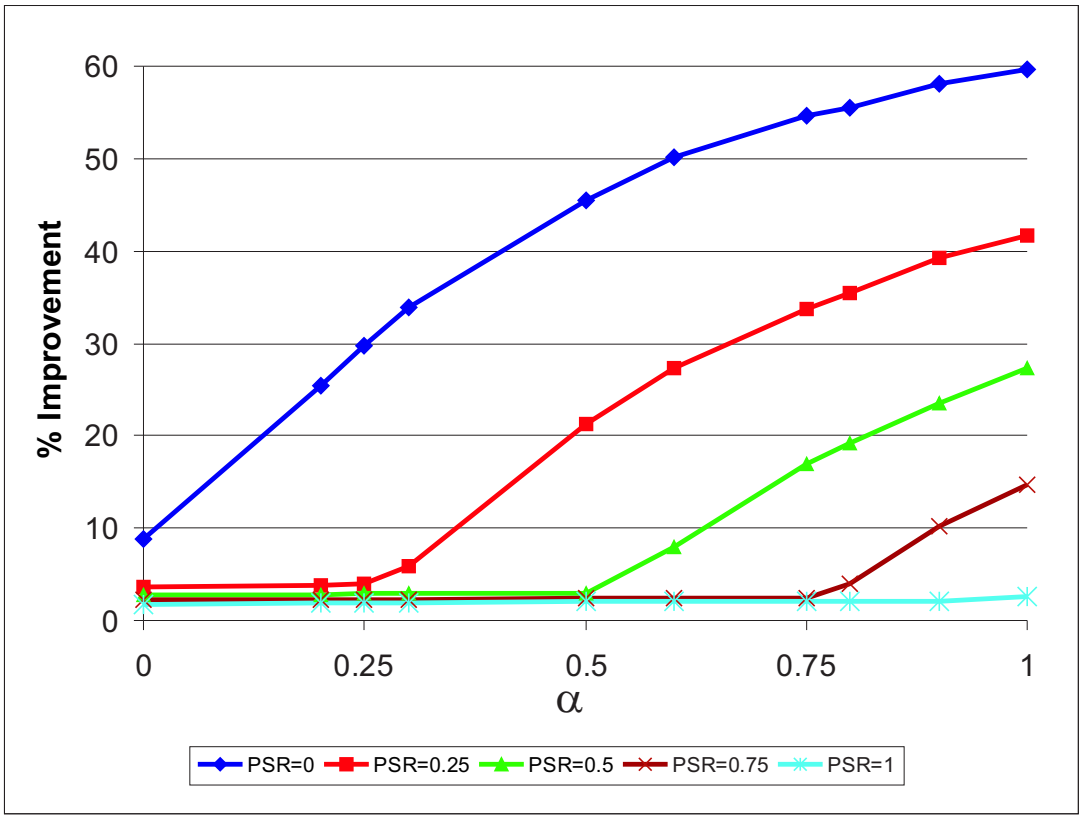



**Figure 16:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for Different Levels of Product Similarity

not have a significant effect on the performance of the suboptimal policy. It can be seen in Figure 17 that as $K$ increases, $V_s^{LB(K)}$ is also observed to increase; Proposition 4 predicts this behavior. However, the percentage difference between $V_s^{LB(4)}$ and $V_s^{LB(1)}$ is only about 0.2%. Thus, the results obtained in the first experiment do not appear to be significantly affected by the chosen value of $K = 2$ for cases where $\alpha$ is approximately equal to or lower than $PSR$. In contrast, when $PSR = 0.25$ the value of $K$ has a significant effect on the performance of the suboptimal policy (see Figure 18), the percentage difference between $V_s^{LB(4)}$ and $V_s^{LB(1)}$ is about 13%. Thus, the results obtained in the first experiment are significantly affected by the chosen value of $K = 2$ for cases where $\alpha$ is significantly larger than $PSR$. Once again suggesting that for computationally tractable values of $K$ the suboptimal policy does not perform very well in systems with high substitutability and low $PSR$ levels.

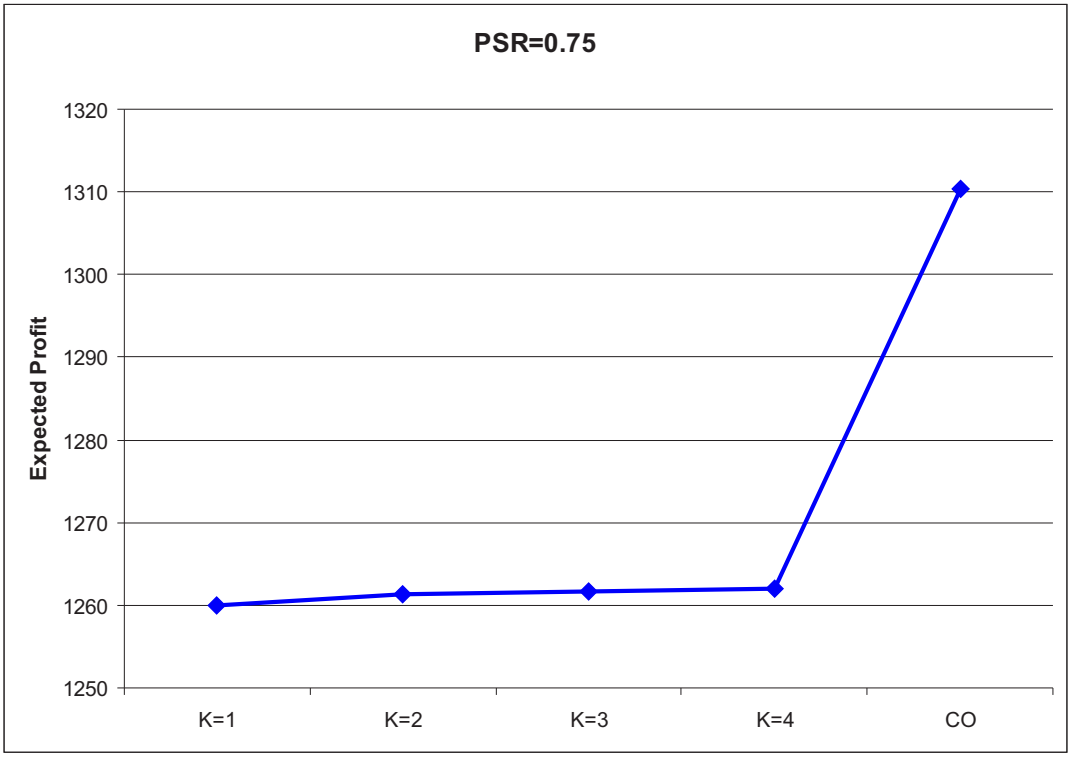



**Figure 17:** Expected Profit of the Sales-Only-Observed Case ($V_s^{LB(K)}$) for Different Levels of $K$ and the Completely-Observed Case $V_O^*$ ($PSR = 0.75$)

The objective of the third experiment is to obtain insights on the effect that the level of substitutability has on the value of improved demand observability. The maximum demand of products type 1 and type 2 are set to be $D_1 = 4$ and $D_2 = 2$ respectively. The discount factor and the parameter $K$ chosen were $\beta = 0.95$ and $K = 2$. The parameter $r$ of the Markovian demand process of product type 1 is set to 1, but the parameter $\zeta$ is varied between scenarios, and is selected from the set $\{0.75, 0.85, 0.95\}$. The per unit ordering cost, selling price and holding cost of products type 1 and type 2 is set to be $c_1 = 145$, $p_1 = 110$, $h_1 = 2.5$, $c_2 = 160$, $p_2 = 200$ and $h_2 = 3$, thus the $PSR = 0.875$.

This experiment suggests, that type 1 demand variability does not have a significant effect on the value of improved demand observability (see Figure 19). It can also be observed in Figure 19 that substitution has a significant effect in the value of improved demand observability. As the substitution likelihood $\alpha$ increases, the value of improved demand observability also increases. Note that as $\alpha$ grows from zero to one, the average value of

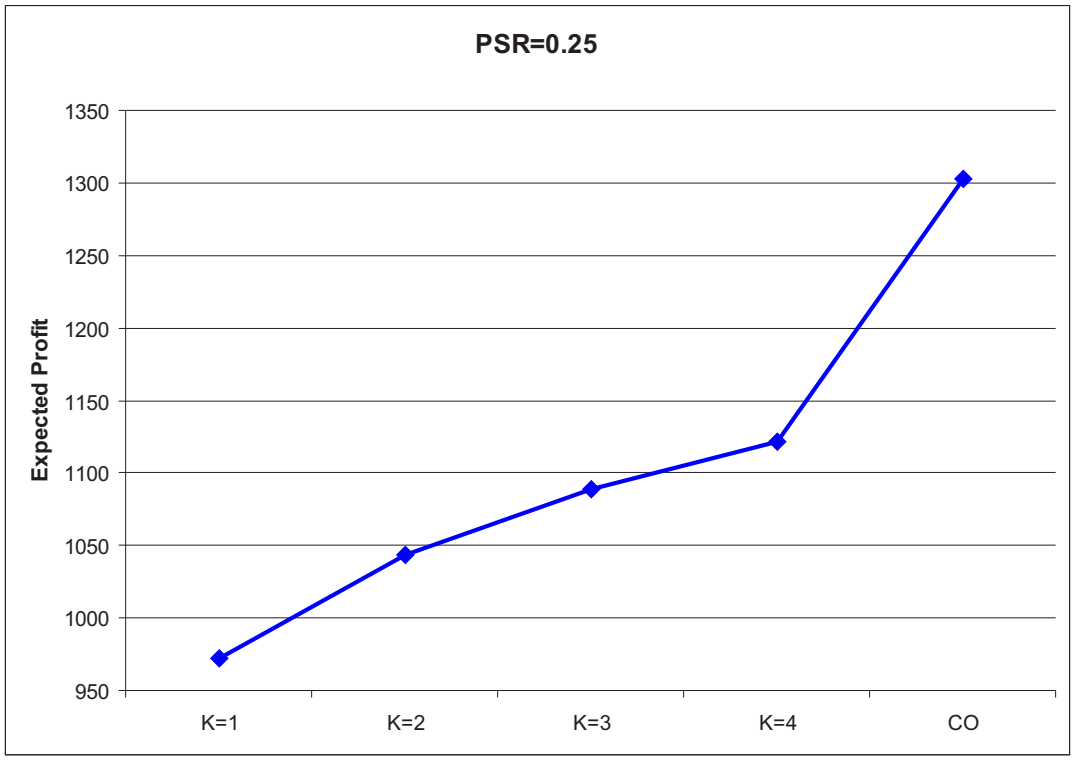



**Figure 18:** Expected Profit of the Sales-Only-Observed Case ($V_s^{LB(K)}$) for Different Levels of $K$ and the Completely-Observed Case $V_O^*$ ($PSR = 0.25$)

observability more than doubles from 4.4% to 10.85%.

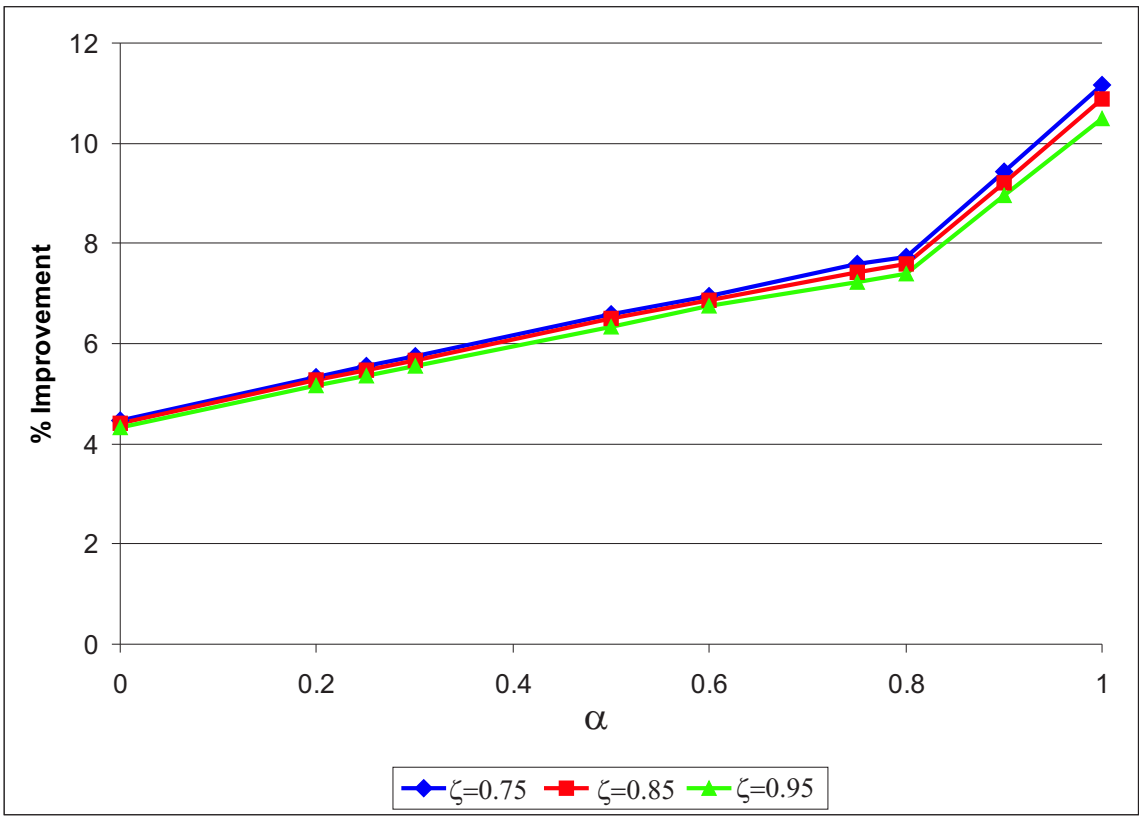



**Figure 19:** Weighted Maximum Percentage Improvement in Profitability Given Demand Observability ($G^{UB(2)}$) for Different Levels of Type 1 Demand Variability

# CHAPTER V

# AN INFINITE HORIZON, TWO-ITEM INVENTORY PROBLEM WITH SUBSTITUTABILITY

## *5.1 Introduction*

We consider the problem of determining an optimal replenishment policy for a two-item, infinite horizon inventory problem. Demand for each item during each period is deterministic. Replenishment decisions occur at each period and replenishment is instantaneous. Profit per item sold, wholesale cost per item, and holding cost per item per period, for each of the two items, are used to construct the single period cost function. The criterion of interest is the infinite horizon, expected total discounted cost criterion.

We allow no backlogging. However, substitutability is allowed. We assume that a customer who wishes to purchase item 1 and finds item 1 stocked out may be willing to purchase item 2 if item 2 is available and that a customer who wishes to purchase item 2 and finds item 2 has stocked out has no interest in purchasing item 1. Substitutability is modeled by the conditional probability $P(d'|d)$, where $d$ is the number of customers who wish to purchase item 1 but find item 1 stocked out, and $d'$ is the number of these customers who wish to purchase item 2. Literature related to inventory control systems with substitution can be found in Section 4.2 of Chapter 4.

This chapter is outlined as follows. In Sections 5.2 and 5.3, we formulate the problem and present preliminary results. These results include the optimality equation for the problem.

Section 5.4 is focused on the single period cost function, $f$, which is a function of the inventory levels of both products. In Section 5.4.1 we present a useful partition generated by $f$ and show that the minimum of $f$ is restricted to one of the elements of this partition. We then determine conditions that guarantee the minimum of $f$ is such that zero replenishment of item 1 is always optimal and examine these bounds in the context of two distributions of substitutability, the uniform and the binomial distributions. We present an algorithm

in Section 5.4.3 that determines the policy that minimizes $f$. We show that for a fixed inventory level for item 1 $f$ is convex in the inventory level of item 2 and give conditions that imply $f$ is convex in the inventory level of item 1 for fixed inventory level for item 2. When these two convexity conditions hold, the algorithm returns an optimal policy that is a generalization of the order-up-to policy associated with the single item special case. We show that this algorithm simplifies significantly when conditions hold for the existence of an optimal zero replenishment policy for item 1.

We present two results in Section 5.5. We first show that greater substitutability will not increase optimal expected cost and use this result to generate upper and lower bounds on optimal expected cost. We then show that a decision rule that minimizes $f$, when applied at every decision epoch over the infinite horizon, is an optimal (myopic) policy for the infinite horizon problem.

## *5.2 Problem Formulation*

Let $x_i(t)$ be the number of items of product $i$ in inventory at the beginning of period $t$, $i = 1, 2$. Based on these inventory levels, the decision maker orders $a_i(t)$ items of product $i$, $i = 1, 2$, which immediately are added to the inventories. Thus, there are $y_i(t) = x_i(t) + a_i(t)$ items of product $i$, $i = 1, 2$, in inventory at the beginning of period $t$.

Let $d_i$ be the demand for item $i$ during each period. Note that if $d_1 > y_1(t)$, then product 1 stocks out during period $t$ and the unmet demand for product 1 is $d = d_1 - y_1(t)$. Of these $d$ customers, let $d_2'(t) \leq d$ be the number of customers willing to purchase product 2, and let $P(d'|d)$ be the probability that $d_2'(t) = d'$.

We assume no backlogging. Thus,

$$x_1(t+1) = \max\{0, y_1(t) - d_1\}$$

$$x_2(t+1) = \max\{0, y_2(t) - d_2 - d_2'(t)\},$$

where $d_2'(t) = d'$ with probability $P(d'|d)$ and $d = \max\{0, d_1 - y_1(t)\}$.

For a single item of product $i$, let $c_i$ be its wholesale cost, $h_i$ be its single period holding cost, and $p_i$ be the profit accrued by selling it. Then, the single period total cost is:

$$\sum_i c_i a_i(t) + \sum_i h_i y_i(t) - p_1 \min\{d_1, y_1(t)\} - p_2 \min\{d_2 + d_2'(t), y_2(t)\}.$$

A (stationary) policy maps the set of inventory levels, $[0, 1, \cdots)^2$, into the set of actions (or replenishment levels), $[0, 1, \cdots)^2$. The problem objective is to determine a policy, called an optimal policy, that minimizes the expected total discounted cost criterion over the infinite horizon. See (Puterman [40], Chapter 6) for further details.

### *5.3 Preliminary Results*

For notational simplicity, we now drop explicit dependence on $t$. The optimality equation for discount factor $\beta < 1$ then becomes:

$$\begin{aligned}\overline{v}(x_1, x_2) \quad = \quad \min_{a_i \geq 0} \Bigg\{ & \sum_i c_i a_i + \sum_i h_i y_i - p_1 \min\{d_1, y_1\} \\ & -p_2 \sum_{d_2'} P(d_2' | \max\{0, d_1 - y_1\}) \min\{d_2 + d_2', y_2\} \\ & +\beta \sum_{d_2'} P(d_2' | \max\{0, d_1 - y_1\}) \overline{v} \left[\max\{0, y_1 - d_1\}, \max\{0, y_2 - d_2 - d_2'\}\right] \Bigg\}.\end{aligned}$$

Replace $a_i$ with $y_i - x_i$, let $v(x_1, x_2) = \overline{v}(x_1, x_2) + \sum_i c_i x_i$, and note that for any constant $k$, $y = \min\{k, y\} + \max\{0, y - k\}$. Straightforward algebraic manipulation then implies that the optimality equation can be re-stated as $v = Hv$, where:

$$[Hv](x_1, x_2) = \min_{y_i \geq x_i} \left\{ [\tilde{H}v](y_1, y_2) \right\}, \tag{16}$$

$$\begin{aligned}[\tilde{H}v](y_1, y_2) \quad = \quad & f(y_1, y_2) + \\ & \beta \sum_{d_2'} P(d_2' | \max\{0, d_1 - y_1\}) v \left[\max\{0, y_1 - d_1\}, \max\{0, y_2 - d_2 - d_2'\}\right],\end{aligned}$$

and

$$f(y_1, y_2) = \sum_i \bar{h}_i y_i - \bar{p}_1 \min\{d_1, y_1\} - \bar{p}_2 \sum_{d_2'} P(d_2' | \max\{0, d_1 - y_1\}) \min\{d_2 + d_2', y_2\},$$

where the sums with respect to $d_2'$ are such that $d_2' = 0, 1, \cdots, \max\{0, d_1 - y_1\}$ and $\bar{h}_i = h_i + (1 - \beta)c_i$ and $\bar{p}_i = p_i - \beta c_i$, $i = 1, 2$.

Throughout, we assume that $\bar{p}_i \geq \bar{h}_i \geq 0$ for $i = 1, 2$.

Results in (Puterman [40], Chapter 6) imply that there exists a unique solution to $v = Hv$; let $v^*$ represent this unique solution. Then, $\overline{v}(x_1, x_2) = v^*(x_1, x_2) - \sum_i c_i x_i$ represents the minimum expected total infinite horizon discounted cost. Furthermore, let $v_0$ be any bounded, real-valued function, and define the sequence $\{v_n\}$ by $v_{n+1} = Hv_n$. Then, $\{v_n\}$ converges to $v^*$ in the sense that $\lim_{n\to\infty} \|v_n - v^*\| = 0$, where $\|\cdot\|$ is the supremum norm. Further, an action selection rule that selects $a_1$ and $a_2$ so as to cause the minimum in Equation (16) to be attained, as a function of $(x_1, x_2)$, is an optimal (stationary) policy.

### *5.4 Structural Properties of $f$*

We now investigate important structural properties of the function $f$. We begin by presenting a useful partition generated by $f$ and show that the minimum of $f$ is restricted to one of the elements of this partition. We then determine conditions that guarantee the minimum of $f$ is such that zero replenishment of item 1 is always optimal and examine these in the context of two distributions of substitutability, the uniform and the binomial distributions. We then present an algorithm that determines a decision rule that minimizes $f$. We show that for a fixed inventory level for item 1, $f$ is convex in the inventory level of item 2 and give conditions that imply $f$ is convex in the inventory level of item 1 for a fixed inventory level for item 2. When these two convexity conditions hold, the algorithm returns an optimal policy that is a generalization of the order-up-to policy associated with the single item special case. We then show that this algorithm simplifies significantly when conditions hold for the existence of an optimal zero replenishment policy for item 1.

### 5.4.1 A Partition of the Inventory Levels

We now present a useful partition generated by $f$ and show that the minimum of $f$ is restricted to one of the elements of this partition. There are five areas within $\{(y_1, y_2), y_i \geq 0\}$ that are of interest:

$$\begin{aligned}
\mathcal{P}(1) &= \{(y_1, y_2) : y_1 \geq d_1, y_2 \geq d_2\} \\
\mathcal{P}(2) &= \{(y_1, y_2) : y_1 \geq d_1, y_2 \leq d_2\} \\
\mathcal{P}(3) &= \{(y_1, y_2) : y_1 \leq d_1, y_2 \leq d_2\} \\
\mathcal{P}(4) &= \{(y_1, y_2) : y_1 \leq d_1, y_2 \geq d_2 + (d_1 - y_1)\} \\
\mathcal{P}(5) &= \{(y_1, y_2) : y_1 \leq d_1, d_2 \leq y_2 \leq d_2 + (d_1 - y_1)\},
\end{aligned}$$

which are depicted in Figure 20. We note that $f$ is:

(i) isotone (monotonically non-decreasing) in $y_1$ and $y_2$ on $\mathcal{P}(1)$

(ii) isotone in $y_1$ and antitone (monotonically non-increasing) in $y_2$ on $\mathcal{P}(2)$

(iii) antitone in $y_1$ and $y_2$ on $\mathcal{P}(3)$

(iv) isotone in $y_2$ on $\mathcal{P}(4)$.

Assume, for a moment, that there is no substitutability, *i.e.*, assume $P(0|d) = 1$ for all $d$. Then, $f$ is identical on both $\mathcal{P}(4)$ and $\mathcal{P}(5)$ and is antitone in $y_1$ and isotone in $y_2$. Thus, for the $P(0|d) = 1$ case, $f(d_1, d_2) \leq f(y_1, y_2)$ for all $(y_1, y_2)$.

Sufficiently strong substitutability, however, can change $f$ on $\mathcal{P}(4)$ from antitone in $y_1$ and isotone in $y_2$ to isotone in both $y_1$ and $y_2$, leaving the structure of $f$ on $\mathcal{P}(1)$, $\mathcal{P}(2)$, and $\mathcal{P}(3)$ unaffected. In such a situation, a point in $\mathcal{P}(5)$ other than $(d_1, d_2)$ might represent a minimum of $f$, as the following example demonstrates.

**Example 1** *Let:* $c_1 = 160$, $c_2 = 200$, $h_1 = 6$, $h_2 = 20$, $p_1 = 200$, $p_2 = 300$, $d_1 = 4$, $d_2 = 2$ *and* $\beta = 0.95$. *The conditional probability distributions* $P(d_2' | \max\{0, y_1 - d_1\})$ *for all*

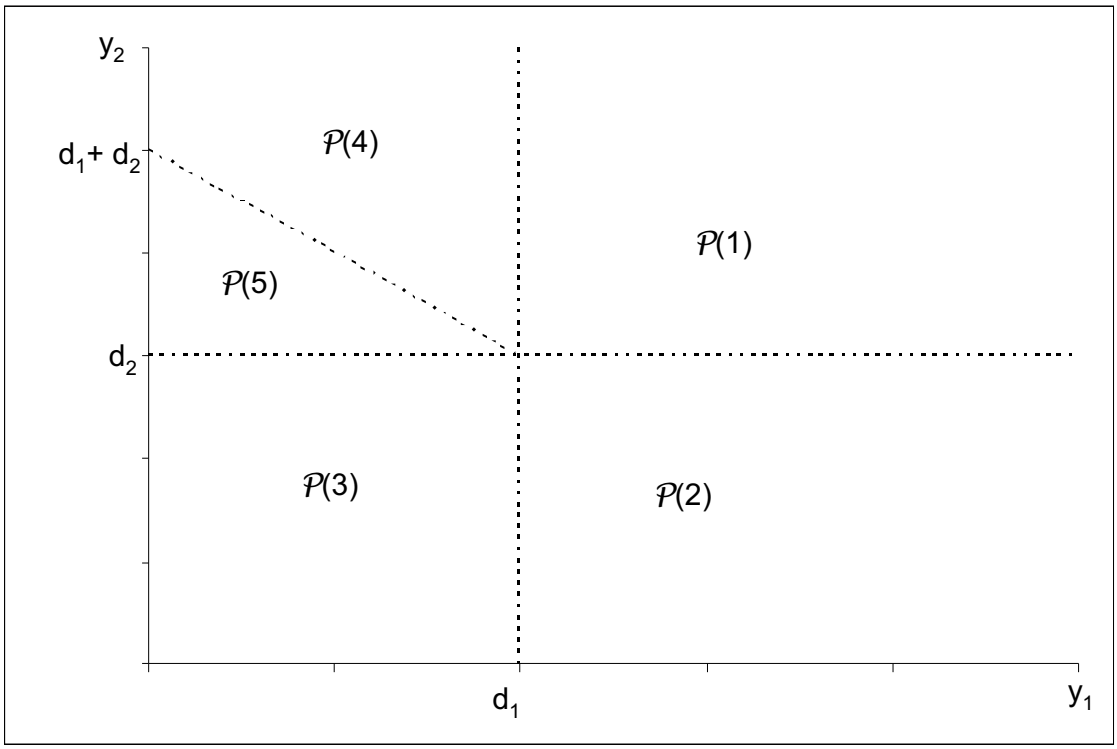



**Figure 20:** Partition of $\{(y_1, y_2) : y_i \geq 0\}$

**Table 3:** $P(d_2'|\max\{0, y_1 - d_1\})$

| | $d_2'$ | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | $> 4$ |
| $P(d_2'\|0)$ | 1 | 0 | 0 | 0 | 0 | 0 |
| $P(d_2'\|1)$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | 0 | 0 | 0 |
| $P(d_2'\|2)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ | 0 | 0 | 0 |
| $P(d_2'\|3)$ | $\frac{1}{8}$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{8}$ | 0 | 0 |
| $P(d_2'\|4)$ | $\frac{3}{16}$ | $\frac{1}{4}$ | $\frac{3}{8}$ | $\frac{1}{8}$ | $\frac{1}{16}$ | 0 |

$0 \leq \max\{0, y_1 - d_1\}$ *are given in Table 3. The point that minimizes* $f$ *is* $(y_1^*, y_2*) = (1, 4)$ *where* $f(1, 4) = -299$ *see Figure* 21.

■

### 5.4.2 Zero Replenishment

We now determine conditions that guarantee the minimum of $f$ is such that zero replenishment of item 1 is always optimal and examine these bounds in the context of two distributions of substitutability, the uniform and the binomial distributions.

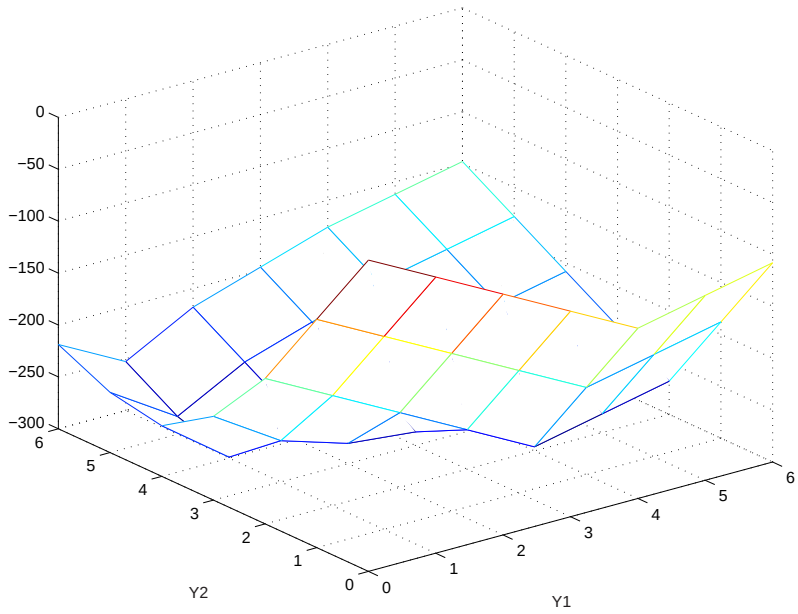



**Figure 21:** $f(y_1, y_2)$

**Theorem 1** *Let* $(x_1, x_2) \in \mathcal{P}(5)$*, and assume that* $f(y_1, y_2) \geq f(y_1 - 1, y_2 + 1)$ *for all* $(y_1, y_2) \in \mathcal{P}(5)$*. Then,*

$$\min_{y_2 \geq x_2} f(x_1, y_2) = \min_{y_i \geq x_i} f(y_1, y_2).$$

**Proof:** Assume $(y_1^*, y_2^*)$ are such that $f(y_1^*, y_2^*) = \min_{y_1 \geq x_1} \min_{y_2 \geq x_2} f(y_1, y_2)$ and $y_1^* \neq x_1$. By assumption, $f(y_1^*, y_2^*) \geq f(y_1^* - 1, y_2^* + 1) \geq \cdots \geq f(x_1, y_2^* + y_1^* - x_1) \geq \min_{y_2 \geq x_2} f(x_1, x_2)$. We remark that if $(y_1^*, y_2^*)$ is in $\mathcal{P}(5)$, then so is $(y_1^* - k, y_2^* + k)$, $k = 1, \cdots, y_1^* - x_1$. Let $y_2'$ be such that $f(x_1, y_2') = \min_{y_2 \geq x_2} f(x_1, y_2)$. If all of the inequalities in the above string of inequalities are equalities, then $(x_1, y_2')$ is also an optimal solution. If any of the inequalities is strict, then we have a contradiction to the claim that $y_1^* \neq x_1$ is optimal.

■

We remark that $f(y_1, y_2) \geq f(y_1 - 1, y_2 + 1)$ for all $(y_1, y_2) \in \mathcal{P}(5)$ implies that we would prefer to replace a unit of item 1 with a unit of item 2. Further, this assumption implies that $f(d_1, d_2) \geq f(0, d_1 + d_2)$ and hence that $f(y_1, y_2)$ is isotone in both $y_1$ and $y_2$ on $\mathcal{P}(4)$.

We now present a condition on the cost structure and on $P$ that guarantees the hypothesis of Theorem 1 holds. Proof is straightforward. Let

$$\begin{aligned}\sigma(d_1,d_2,y_1,y_2) &= \sum_{d_2'=0}^{y_2-d_2+1} P(d_2'|d_1-y_1+1)(d_2+d_2') \\ &+ \sum_{d_2'=y_2-d_2+2}^{d_1-y_1+1} P(d_2'|d_1-y_1+1)(y_2+1) \\ &- \left[\sum_{d_2'=0}^{y_2-d_2} P(d_2'|d_1-y_1)(d_2+d_2') + \sum_{d_2'=y_2-d_2+1}^{d_1-y_1} P(d_2'|d_1-y_1)y_2\right].\end{aligned}$$

**Lemma 2** *Assume* $(y_1,y_2)\in\mathcal{P}(5)$. *Then,* $f(y_1,y_2)\geq f(y_1-1,y_2+1)$ *if and only if*

$$\sigma(d_1,d_2,y_1,y_2)\bar{p}_2-\bar{h}_2\geq\bar{p}_1-\bar{h}_1.$$

We now consider two conditional distributions, the uniform and the binomial, and determine their values of $\sigma$. We then determine lower bounds on $\sigma$, which are sufficient conditions for zero replenishment of product 1 to be optimal. We begin by putting $\sigma$ into a more useful form.

Let $m$ and $n$, $m\leq n$, be such that $y_2=d_2+m$ and $y_1=d_1-n$. Then,

$$\begin{aligned}\sigma(d_1,d_2,d_1-n,d_2+m) &= \sum_{k=0}^{m+1}(k+d_2)P(k|n+1)+\sum_{k=m+2}^{n+1}(d_2+m+1)P(k|n+1) \\ &- \sum_{k=0}^{m}(k+d_2)P(k|n)-\sum_{k=m+1}^{n}(d_2+m)P(k|n).\end{aligned}$$

Since $\sum_{k=0}^{n}P(k|n)=\sum_{k=0}^{n+1}P(k|n+1)=1$, dependence on $d_2$ sums to zero on the right hand side of the above equation. Note also that,

$$\sum_{k=0}^{m}kP(k|n)+m\sum_{k=m+1}^{n}P(k|n) = E(n)-\sum_{k=m+1}^{n}(k-m)P(k|n),$$

where $E(n)=\sum_{k=0}^{n}kP(k|n)$. Thus, $\sigma(n,m)=\sigma(d_1,d_2,d_1-n,d_2+m)$ can be written as:

$$\sigma(n,m)=E(n+1)-E(n)+\sum_{k=m+1}^{n}(k-m)[P(k|n)-P(k+1|n+1)].$$

Let $P(k|n) = \frac{1}{(n+1)}$ for all $k = 0, 1, \cdots, n$, which we call the uniform distribution. Thus, if $n$ customers who want to purchase product 1 find that product 1 has stocked out, then $k$ of these customers will want to purchase product 2 with probability $\frac{1}{(n+1)}$, for all $k = 0, 1, \cdots, n$.

Proof of the following result is straightforward.

**Lemma 3** *For $m = 0, 1, \ldots, n$, $\sigma(n, m) = \frac{1}{2} + \frac{(n-m)(n-m+1)}{2(n+1)(n+2)}$, where $P(k|n) = \frac{1}{(n+1)}$, $k = 0, 1, \cdots, n$.*

Thus, $\sigma(n, m) \geq \frac{1}{2}$ for the uniform distribution for all $m$ and $n$, $m \leq n$, implying the following result.

**Lemma 4** *For the uniform distribution, if*

$$\frac{\bar{p}_2}{2} - \bar{h}_2 \geq \bar{p}_1 - \bar{h}_1,$$

*then for all $(x_1, x_2) \in \mathcal{P}(5)$,*

$$\min_{y_2 \geq x_2} f(x_1, y_2) = \min_{y_i \geq x_i} f(y_1, y_2).$$

Let $P(k|n) = \binom{n}{k}\rho^k(1-\rho)^{n-k}$ for all $k = 0, 1, \cdots, n$, which we call the binomial distribution. We interpret $\rho$ to be the probability that any customer who wants to purchase product 1 and finds product 1 stocked out will want to purchase product 2. The binomial distribution assumes that all such customers act independently.

Proof of the following preliminary result is due to straightforward induction arguments and algebraic manipulation.

**Lemma 5** *Let $P(k|n) = \binom{n}{k}\rho^k(1-\rho)^{n-k}$ for $\rho \in [0, 1]$. Then, $E(n) = \rho n$ and:*

*(i)* $\sum_{k=m}^{n}[P(k|n) - P(k+1|n+1)] = (1-\rho)P(m|n)$

*(ii)* $\sum_{k=m+1}^{n}(k-m)[P(k|n) - P(k+1|n+1)] = (1-\rho)\sum_{k=m+1}^{n} P(k|n)$

*(iii)* $\sum_{k=m+1}^{n}(k-m)P(k|n) - \sum_{k=m+1}^{n+1}(k-m)P(k|n+1) =$
$(1-\rho)\sum_{k=m+1}^{n} P(k|n) - \sum_{k=m+1}^{n+1} P(k|n+1)$

*(iv)* $\sum_{k=m+1}^{n} P(k|n) - \sum_{k=m+1}^{n+1} P(k|n+1) = (1-\rho)P(m+1|n) - P(m+1|n+1)$.

Proof of the next result follows from Lemma 5.

**Lemma 6** *For* $m = 0, 1, \cdots, n$, $\sigma(n,m) = \rho + (1-\rho)\sum_{k=m+1}^{n} P(k|n)$, *where* $P(k|n) = \binom{n}{k}\rho^k(1-\rho)^{n-k}$, $k = 0, 1, \cdots, n$.

Thus, $\sigma(n,m) \geq \rho$ for the binomial distribution for all $m = 0, 1, \cdots, n$, implying the following result.

**Lemma 7** *For the binomial distribution, if*

$$\rho\bar{p}_2 - \bar{h}_2 \geq \bar{p}_1 - \bar{h}_1,$$

*then for all* $(x_1, x_2) \in \mathcal{P}(5)$,

$$\min_{y_2 \geq x_2} f(x_1, y_2) = \min_{y_i \geq x_i} f(y_1, y_2).$$

### 5.4.3 Optimal Policy Structure

We observe that $\mathcal{P}(5)$ contains $N = \frac{(d_1+1)(d_1+2)}{2}$ points. We order these points as follows: $f(y^n) \leq f(y^{n+1})$. Define $\mathcal{P}^n$ as follows, where for each $n$, $\mathcal{P}^n \subseteq \mathcal{P} = \mathcal{P}(3) \bigcup \mathcal{P}(5)$:

0. Let $\mathcal{P}^1 = \{x : x \leq y^1\}$; set $n = 1$.

1. If $y^{n+1} \in \bigcup_{m=1}^{n} \mathcal{P}^m$, then set $\mathcal{P}^{n+1} = \emptyset$. Otherwise, let $\mathcal{P}^{n+1} = \{x \in \mathcal{P} \sim \bigcup_{m=1}^{n} \mathcal{P}^m : x \leq y^{n+1}\}$.

2. Set $n = n + 1$; go to 1.

**Example 2** *Let:* $c_1 = 160$, $c_2 = 200$, $h_1 = 6$, $h_2 = 20$, $p_1 = 200$, $p_2 = 300$, $d_1 = 4$, $d_2 = 2$, $\beta = 0.95$ *and* $P(d_2'|d) = \binom{d}{d_2'}\rho^{d_2'}(1-\rho)^{d-d_2'}$ *where* $\rho = 0.5$. *There are* 15 *points in* $\mathcal{P}(5)$, *see Figure 22. The sets* $\mathcal{P}^n$ *are shown in Figure 23.* $\mathcal{P}^5 = \mathcal{P}^6 = \mathcal{P}^9 = \mathcal{P}^{10} = \mathcal{P}^{12} = \mathcal{P}^{13} = \mathcal{P}^{14} = \mathcal{P}^{15} = \emptyset$.

■

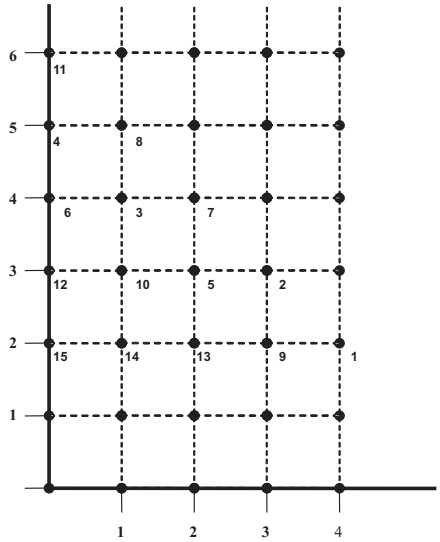



**Figure 22:** Ordered Points in $\mathcal{P}(5)$

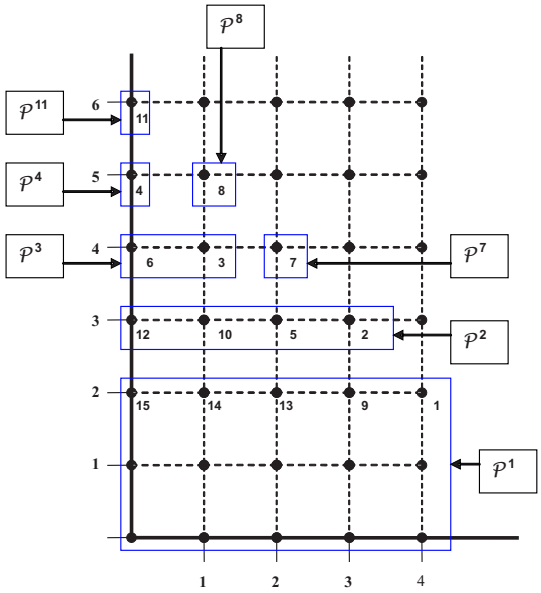



**Figure 23:** Sets $\mathcal{P}^n$

Define $g(x) = \min\{f(y) : x \leq y\}$ and $\delta(x) = \text{argmin}\{f(y) : x \leq y\}$, and note that:

(i) if $x \in \mathcal{P}^n$, then $g(x) = f(y^n)$ and $\delta(x)$ is "order up to $y^n$", $n = 1, \cdots, N$,

(ii) if $x \in \mathcal{P}(1)$ then $g(x) = f(x)$ and $\delta(x)$ is "do not order",

(iii) if $x \in \mathcal{P}(2)$, then $g(x_1, x_2) = f(x_1, d_2)$ and $\delta(x)$ is "do not order item 1 but order item 2 up to $d_2$",

(iv) if $x \in \mathcal{P}(4)$, then

(a) $g(x) = f(x)$ and $\delta(x)$ is "do not order" if $f(0, d_1 + d_2) \leq f(d_1, d_1 + d_2)$

(b) $g(x_1, x_2) = f(d_1, x_2)$ and $\delta(x)$ is "order item 1 up to $d_1$ and do not order item 2" otherwise.

We remark that if $\mathcal{P}$ is a singleton, then "order up to $y^n$" is equivalent to "do not order". We elaborate on this remark by showing $f(y_1, y_2)$ is convex in $y_2$ for fixed $y_1$ on $\mathcal{P}(5)$, presenting sufficient conditions for $f(y_1, y_2)$ to be convex in $y_1$ on $[0, d_1]$ for fixed $y_2 \in [d_2, d_1 + d_2]$, and then using these two convexity results to simplify the algorithm once a $\mathcal{P}^n$ is determined that is a singleton.

**Lemma 8** *For each $y_1$, $f(y_1, y_2)$ is convex in $y_2$.*

**Proof:** The result is clearly true for $y_1 \geq d_1$. Assume $y_1 \leq d_1$. It is sufficient to show that

$$f(y_1, y_2 + 2) - f(y_1, y_2 + 1) \geq f(y_1, y_2 + 1) - f(y_1, y_2). \tag{17}$$

This inequality clearly holds when $y_2 + 2 \leq d_2$ or $d_2 + (d_1 - y_1) \leq y_2$. There are three other cases:

(i) $y_2 = d_2 + (d_1 - y_1) - 1$,

(ii) $y_2 = d_2 - 1$,

(iii) $d_2 \leq y_2 \leq d_2 + (d_1 - y_1) - 2$.

We now consider case $(i)$; cases $(ii)$ and $(iii)$ follow in a similar fashion.

Let $n = d_1 - y_1$. Then, it is easily shown that:

$$f(y_1, d_2 + n - 1) = (\bar{h}_1 - \bar{p}_1)y_1 + \bar{h}_2(d_2 + n - 1) - \bar{p}_2[d_2 + E(n) - P(n|n)],$$

$$f(y_1, d_2 + n) = (\bar{h}_1 - \bar{p}_1)y_1 + \bar{h}_2(d_2 + n) - \bar{p}_2[d_2 + E(n)],$$

and

$$f(y_1, d_2 + n + 1) = (\bar{h}_1 - \bar{p}_1)y_1 + \bar{h}_2(d_2 + n + 1) - \bar{p}_2[d_2 + E(n)].$$

Algebraic manipulation then shows that the inequality in (17) is equivalent to $\bar{p}_2 P(n|n)(d_2 + n) \geq 0$.

■

We now present conditions that imply $f(y_1, y_2)$ is convex in $y_1$ on $[0, d_1]$ for fixed $y_2$. Let $m$ and $n$ be such that $y_1 = d_1 - n$ and $y_2 = d_2 + m$. Thus, for $y \in \mathcal{P}(5)$, $0 \leq m \leq n \leq d_1$ and

$$f(d_1 - n, d_2 + m) = K + (\bar{p}_1 - \bar{h}_1)n + \bar{h}_2 m - \bar{p}_2 F(n, m),$$

where:

$$K = \sum_i d_i(\bar{h}_i - \bar{p}_i),$$

and

$$\begin{aligned} F(n,m) &= \sum_{k=0}^{m} kP(k|n) + m \sum_{k=m+1}^{n} P(k|n) \\ &= E(n) - \sum_{k=m+1}^{n} (k-m)P(k|n). \end{aligned}$$

**Lemma 9** *Assume $E(n+2) - E(n+1) = E(n+1) - E(n)$ for all $n$ such that $d_1 - 2 \geq n \geq 0$. Assume also that for all $n$ and $m$ such that $d_1 - 2 \geq n \geq m \geq 0$,*

$$\sum_{k=m+1}^{n+1}(k-m)P(k|n+1) - \sum_{k=m+1}^{n+2}(k-m)P(k|n+2) \leq \sum_{k=m+1}^{n}(k-m)P(k|n) - \sum_{k=m+1}^{n+1}(k-m)P(k|n+1).$$

*Then, for fixed $y_2 \in [d_2, d_1 + d_2]$, $f(y_1, y_2)$ is convex in $y_1$ on $[0, d_1]$.*

**Proof:** It is sufficient to show:

(i) for all $n$ and $m$ such that $d_1 \geq n \geq m \geq 0$,

$$f(n+2, m) - f(n+1, m) \geq f(n+1, m) - f(n, m)$$

(ii) for all $m$ such that $d_1 - 1 \geq m \geq 0$,

$$f(m+1,m) - f(m,m) \geq f(m,m) - f(d_1 - m + 1, d_2 + m),$$

where

$$f(d_1 - m + 1, d_2 + m) = K + (\bar{p}_1 - \bar{h}_1)(m-1) + \bar{h}_2 m - \bar{p}_2 E(m-1).$$

Case (i) follows by straightforward algebraic manipulation. Case (ii) follows from the fact that $P(m+1|m+1) \geq 0$.

■

We remark that it is straightforward to show that both the uniform distribution and the binomial distribution (using Lemma 5) satisfy the conditions of Lemma 9.

**Lemma 10** *Assume* $x \in \mathcal{P}(5)$ *and* $\min\{f(y) : x \leq y\} = f(x)$. *Then,*

*(i)* $f(x_1, x_2) \leq f(x_1 + 1, x_2)$ *implies* $\min\{f(y) : x_1' \leq y_1, x_2 \leq y_2\} = f(x_1', x_2)$ *for all* $x_1' \geq x_1$, *assuming* $f(y_1, y_2)$ *is convex in* $y_1$ *for any given* $y_2$

*(ii)* $f(x_1, x_2) \leq f(x_1, x_2 + 1)$ *implies* $\min\{f(y) : x_1 \leq y_1, x_2' \leq y_2\} = f(x_1, x_2')$ *for all* $x_2' \geq x_2$.

We now examine the algorithm under the assumption that $f(y_1, y_2) \geq f(y_1 - 1, y_2 + 1)$ in $\mathcal{P}(5)$. Let $R(y_1)$ be such that $f(y_1, R(y_1)) \leq f(y_1, y_2)$ for all $y_2$. We note that $f(y_1, R(y_1))$ is isotone since $f(y_1 + 1, R(y_1 + 1)) \geq f(y_1, R(y_1 + 1) + 1) \geq f(y_1, R(y_1))$, for all $y_1$. We now present sufficient conditions for $R(y_1)$ to be antitone.

**Lemma 11** *Assume for* $n \leq n'$

$$\sum_{k \geq m} P(k|n) \leq \sum_{k \geq m} P(k|n'),$$

*for all* $m \leq n$. *Then* $R(y_1)$ *is antitone.*

**Proof:** The result holds for $y_1 \geq d_1$; assume $y_1 \leq d_1$. It is shown in (Puterman [40]; Lemma 4.7.1, p. 104) that $R(y_1)$ is antitone if $f$ is superadditive; *i.e.*, $y_1 \leq y_1'$ and $y_2 \leq y_2'$ imply

$$f(y_1, y_2) - f(y_1', y_2) \geq f(y_1, y_2') - f(y_1', y_2'). \tag{18}$$

Without loss of generality, let $y_1 = d_1 - n$, $y_1' = d_1 - n'$, $n' = n-1$, $y_2 = d_2 + m$, $y_2' = d_2 + m'$, and $m' = m + 1$.

Algebraic manipulation indicates that Equation (18) is equivalent to:

$$\sum_{k=m+1}^{n} (k-m)P(k|n) - \sum_{k=m+1}^{n-1} (k-m)P(k|n-1) \geq \sum_{k=m+2}^{n} (k-m-1)P(k|n) - \sum_{k=m+2}^{n-1} (k-m-1)P(k|n-1),$$

and hence

$$\sum_{k=m+1}^{n} P(k|n) \geq \sum_{k=m+1}^{n-1} P(k|n-1).$$

■

It is easily shown that both the uniform and the binomial (where the right hand side of the above inequality equals $(1-\rho)P(m+1|n-1)$; see Lemma 5) distributions satisfy this inequality and hence $f$ is superadditive for both distributions.

Assume that $R(y_1)$ is antitone and $f(y_1, R(y_1))$ is isotone. Then the algorithm reduces to:

(i) if $(x_1, x_2)$ is such that $x_2 \leq R(x_1)$, then $g(x_1, x_2) = f(x_1, R(x_1))$ and $\delta(x_1, x_2) =$ "do not order item 1 and order up to $R(x_1)$ of item 2".

(ii) if $(x_1, x_2)$ is such that $x_2 \geq R(x_1)$, then $g(x_1, x_2) = f(x_1, x_2)$ and $\delta(x_1, x_2) =$ "order neither item 1 nor item 2".

### *5.5 Infinite Horizon Case*

We now present two results for the infinite horizon case. The first result is that increased substitutability will never increase optimal expected discounted cost. The second result is

that an optimal decision rule for the single-stage case, when applied at every decision epoch over the infinite horizon, is an optimal policy for the infinite horizon case.

### 5.5.1 Substitutability and Bounds

We now present a definition of increased substitutability. We say $P'$ has increased substitutability, relative to $P$, if and only if for each $d$ and each $k = 0, \cdots, d$,

$$\sum_{d' \geq k} P(d'|d) \leq \sum_{d' \geq k} P'(d'|d).$$

We remark that this concept is related to conditions that imply the existence of optimal monotone policies, as presented in (Puterman [40]; Chapter4, Section 7).

Increasing the parameter $\rho$ in the binomial distribution leads to increased substitutability, as we now show.

**Lemma 12** *Let $P(k|n) = \binom{n}{k}\rho^k(1-\rho)^{n-k}$, and let $P'(k|n)$ equal $P(k|n)$ with $\rho$ replaced by $\rho'$. Then $\rho \leq \rho'$ implies $P'$ has increased substitutability, relative to $P$.*

**Proof:** The cumulative distribution function of the binomial distribution can be expressed in terms of the incomplete beta function as follows:

$$\sum_{k=0}^{m-1} \binom{d}{k} \rho^k (1-\rho)^{d-k} = I_{1-\rho}(d-k+1, k)$$

where

$$I_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1}dt}{\int_0^\infty t^{a-1}(1+t)^{-(a+b)}dt}$$

it follows then that $I_{1-\rho}(d-k+1,k) \geq I_{1-\rho'}(d-k+1,k)$.

■

Let the operators $\tilde{H}'$ and $H'$ be defined identically to $\tilde{H}$ and $H$ with $P$ replaced by $P'$, and assume $v$ and $v'$ are the fixed points of the operators $H$ and $H'$, respectively. We now present our main result for this section.

**Theorem 2** *Assume $P'$ has increased substitutability, relative to $P$. Then, $v \geq v'$.*

**Proof:** It is sufficient to show that $[\tilde{H}v](y_1, y_2) \geq [\tilde{H}'v](y_1, y_2)$ for any $v(x_1, x_2)$ that is isotone and convex in $x_2$ for all $x_1$. We remark that for any $v$, $Hv$ is isotone and hence $H$ has an isotone fixed point. Thus, it is without loss of generality that we assume $v$ is isotone. Referring to the definition of $[\tilde{H}v](y_1, y_2)$, we note that $-\min\{d_2 + d_2', y_2\}$ and $\max\{0, y_2 - d_2 - d_2'\}$ are both antitone in $d_2'$ and hence

$$\bar{p}_2(-\min\{d_2 + d_2', y_2\}) + \beta v[\max\{0, y_1 - d_1\}, \max\{0, y_2 - d_2 - d_2'\}]$$

is antitone in $d_2'$. The sufficient condition and hence the result follows from Lemma 4.7.2 (Puterman [40], p. 106).

■

Let $P^m(d'|d) = 1$ if and only if $d' = 0$. Note that any $P$ has increased substitutability, relative to $P^m$. Let $P^M(d'|d) = 1$ if and only if $d' = d$. Then, $P^M$ has increased substitutability, relative to any $P$. Let the operators $H^m$ ($H^M$) be defined as the operator $H$ with $P^m$ ($P^M$) replacing $P$, and let $v^m$ ($v^M$) be the fixed point of $H^m$ ($H^M$). The next result then follows directly from Theorem 2.

**Corollary 3** $v^m \geq v \geq v^M$.

We remark that $v^m$ is relatively easy to determine. Proof of the next result follows directly from the optimality equation.

**Theorem 3** *There exit functions $v_1^m(x_1)$ and $v_2^m(x_2)$ such that $v^m(x_1, x_2) = v_1^m(x_1) + v_2^m(x_2)$, where $v_i^m(x_i)$ is the fixed point of the operator $H_i^m$, $[H_i^m v](x_i) = \min_{y_i \geq x_i}[\tilde{H}_i^m v](y_i)$, and*

$$[\tilde{H}_i^m v](y_i) = \bar{h}_i y_i - \bar{p}_i \sum_{d_i} P_i(d_i) \min\{d_i, y_i\} + \beta v[\max\{0, y_i - d_i\}].$$

### 5.5.2 Myopic Optimal Policies

Given $f$, we have determined the collection of sets $\{\mathcal{P}^n\}$ and vectors $\{y^n\}$ that characterize an optimal decision rule for the single stage problem. We now show that this decision rule, used at each decision epoch for the infinite horizon problem, is optimal. We present this result, following preliminary definitions.

Let $V^*$ be the set of all real-valued functions on $\{(y_1, y_2) : y \geq 0\}$ defined as follows: $v \in V^*$ if and only if $v$ is isotone and constant for all $y \leq y^1$.

Let $V^{**}$ be the set of all real-valued functions on $\{(y_1, y_2) : y \geq 0\}$ defined as follows: $v \in V^{**}$ if and only if $v$ is

(i) isotone on $\mathcal{P}(1)$

(ii) isotone in $y_1$ and antitone in $y_2$ on $\mathcal{P}(2)$

(iii) antitone on $\mathcal{P}(3)$

(iv) isotone in $y_2$ on $\mathcal{P}(4)$

(v) $v(y^n) \leq v(y^{n+1}), n = 1, \cdots, N-1$.

We note that $f \in V^{**}$ and $g \in V^*$.

We now show that there exists an optimal myopic policy.

**Theorem 4** *Assuming $d_1 \leq y_2^1$,*

*(i) if $v \in V^*$, then $\tilde{H}v \in V^{**}$*

*(ii) if $\tilde{H}v \in V^{**}$, then $Hv \in V^*$*

*(iii) if $v^*$ is the fixed point of $H$, then $v^* \in V^*$*

*(iv) $argmin\{f(y) : x \leq y\} = argmin\{[\tilde{H}v^*](y) : x \leq y\}$ and hence there exists an optimal policy that is myopic.*

**Proof:** We note that (ii) holds by the construction of $V^{**}$ and $V^*$. Regarding (iii), if $v \in V^*$ implies $Hv \in V^*$, then $v^* \in V^*$ since $v^n$ converges to $v^*$, $v^{n+1} = Hv^n$, where we can choose $v^0 \in V^*$. We note (iv) holds if (ii) and (iii) hold.

We now show that (i) holds. Note that for $y \in \mathcal{P}(5)$,

$$v[\max\{0, y_1 - d_1\}, \max\{0, y_2 - d_2 - k\}] = \begin{cases} v(0, y_2 - d_2 - k) & \text{for } k \leq y_2 - d_2 \\ v(0,0) & \text{for } k \geq y_2 - d_2 \end{cases}.$$

Now, $y_2 - d_2 - k \leq y_2 - d_2 \leq d_1 - y_1 \leq y_2^1 - y_1 \leq y_2^1$. By definition, $v(0, z) = v(y^1)$ for all $z \leq y_2^1$. Thus, $[\tilde{H}v](y) = f(y) + \beta v(y^1)$. Clearly, $v \in V^{**}$ implies $v + c \in V^{**}$ for any constant $c$. Hence $\tilde{H}v \in V^{**}$.

■

# CHAPTER VI

# CONCLUSIONS AND FUTURE RESEARCH

This dissertation investigates the implications of having inaccurate observations of the random realizations of the stochastic process when making decisions under uncertainty in the context of inventory theory. Additionally, the thesis investigates a fundamental problem in inventory management for substitutable products that arose during the study.

Chapter 2 studies the relationship between observation quality and system performance when using zero-memory policies. For zero-memory policies, conditions that imply that improved observation quality (i) will improved system performance, (ii) will degrade system performance, or (iii) will not affect system performance are presented. A computational study of the use of zero-memory policies in a periodic review single item inventory control system with inaccurate counts is presented. Numerical results suggests that inventory systems with high holding cost levels tend to benefit more form improved inventory counts than inventory systems with lower holding cost levels. A natural extension for future research is the behavior of finite-memory policies in this context.

Chapter 3 investigates the maximum value of improving demand observability for periodic-review, single-product inventory systems with unobserved lost sales and Markovian demand. A partially observed Markov decision process model for this system is developed. An algorithm for determining an optimal policy and three computationally attractive heuristics based on a sub-optimal design are presented. A methodology based on the analysis of two extreme cases of the model is used to bound the value of improving observability. A computational study demonstrates the technique, and shows that the bound on profitability gain varies from 2% to over 30% depending on problem characteristics. The sub-optimal design presented in this chapter may overestimate the value of improved demand observability when the inventory holding cost rate is high, since the decision of ordering the maximum possible demand after a certain number of consecutive stockouts may lead to high holding

costs. An interesting future research topic is to consider alternative suboptimal algorithms that may mitigate this effect by slowly increasing the ordering quantity each period a stock-out is observed, up to a maximum value.

Chapter 4 presents an extension of the bounding methodology proposed in Chapter 3 for a two item inventory control system with one way demand substitution. The proposed bounding technique is observed to work fairly well for systems with similar products; however, in the case of high substitutability and high product dissimilarity, the proposed technique appears to overestimate the value of improved demand observability. Numerical results suggest that systems with higher levels of substitution tend to benefit more from improved demand observability than systems with lower levels of substitution. In future research, it would be interesting to consider the effect of improved demand observability in systems with two way substitution.

Chapter 5 presents a two-item inventory system with one way substitution. The system presented assumes deterministic demand, no backlogging, periodic instantaneous replenishment and stochastic substitution. It is shown that a decision rule that minimizes the single period cost function, when applied at every decision epoch over the infinite horizon, is an optimal (myopic) policy for the infinite horizon problem. An in-depth examination of the single period cost function is therefore presented. An algorithm for determining an order-up-to decision rule that minimizes the single period cost function is developed and conditions that imply that the single period cost function is convex in both inventory levels are determined in order to reduce the computational demand of the algorithm. A definition of increased substitutability is presented, and it is shown that increased substitutability never increases optimal expected total discounted cost. The problem presented in this chapter assumes stochastic one way substitution and deterministic demand. Future research could consider systems that faced not only stochastic substitution but also stochastic per period demand.

# REFERENCES


[1] Aberdeen, D. A., *Policy-Gradient Algorithms for Partially Observable Markov Decision Processes*. PhD thesis, The Australian National University, 2003.

[2] Agrawal, N. and Smith, S. A., "Estimating negative binomial demand for retail inventory management with unobservable lost sales," *Naval Research Logistics Quarterly*, vol. 43, pp. 839–861, 1996.

[3] Astrom, K. J., "Optimal control of Markov decision processes with incomplete state estimation," *Journal of Mathematical Analysis and Applications*, vol. 10, 1965.

[4] Azoury, K., "Bayes solution to dynamic inventory models under unknown demand distribution," *Management Science*, vol. 31, pp. 1150–1160, 1985.

[5] Bassok, Y., Anupindi, R., and Akella, R., "Single-period multiproduct inventory models with substitution," *Operations Research*, vol. 47, pp. 632–642, 1999.

[6] Bensoussan, A., Cakanyildirim, M., and Sethi, S. P., "Partially observable inventory systems: the case of zero balance walk," *SIAM Journal of Control and Optimization*, vol. 46, pp. 176–204, 2007.

[7] Bonet, B., "An optimal grid-based algorithm for partially observable Markov decision processes," *In 19th International Conference on Machine Learning, Sydney, Australia*, 2002.

[8] Braden, D. and Freimer, M., "Informational dynamics of censored observations," *Management Science*, vol. 37, pp. 1390–1404, 1991.

[9] Brafman, R. I., "A heuristic variable grid solution method for POMDPs," *In Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI '97)*, 1997.

[10] Cassandra, A., *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes*. PhD thesis, Brown University, 1998.

[11] Cheng, F. and Sethi, S. P., "Optimality of state-dependent (s,S) policies in inventory models with Markov-modulated demand and lost sales," *Production and Operations Management*, vol. 8, pp. 183–192, 1999.

[12] DeHoratios, N., Mersereau, A., and Schrage, L., "Retail inventory management when records are inaccurate," *Graduate School of Business. University of Chicago Working Paper*, 2005.

[13] Ding, X., Puterman, M. L., and Bisi, A., "The censored newsvendor and the optimal acquisition of information," *Operations Research*, vol. 50, pp. 517–527, 2002.

[14] Ernst, R. and Kouvelis, P., "The effects of selling package goods on inventory decisions," *Management Science*, vol. 45, pp. 1142–1155, 1999.

[15] Fisher, M., Raman, A., and McClelland, A., "Rocket science retailing is almost here – are you ready?," *Harvard Business Review*, vol. 74, pp. 115–124, 2000.

[16] Godrey, G. and Powell, W., "An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution," *Management Science*, vol. 47, pp. 1101–1112, 2001.

[17] Hansen, E. A., "Solving POMDPs by searching in policy space," *In The Eighth Conference on Uncertainty in Artificial Intelligence, Madison, WI*, 1998.

[18] Hauskrecht, M., "Value-function approximations for partially observable Markov decision processes," *Journal of Artificial Intelligence Research*, vol. 13, pp. 33–94, 2000.

[19] Iglehart, D., "The dynamic inventory problem with unknown demand distribution," *Management Science*, vol. 10, pp. 429–440, 1964.

[20] Jaakkola, T., Singh, S. P., and Jordan, M. I., "Reinforcement learning algorithm for partially observable Markov decision problems. the MIT press," *In Advances in Neural Information Processing Systems*, vol. 7, pp. 345–352, 1995.

[21] Kang, Y. and Gershwin, S. B., "Information inaccuracy in inventory systems: Stock loss and stockout," *IIE Transactions*, vol. 37, pp. 843– 859, 2005.

[22] Karlin, S., "Dynamic inventory policy with varying stochastic demands," *Management Science*, vol. 6, pp. 231–258, 1960.

[23] Kok, A. G. and Shang, K., "Inspection and replenishment policies for systems with record inaccuracy," *Fuqua School of Business. Duke University Working Paper*, 2006.

[24] Lariviere, M. A. and Porteus, E. L., "Stalking information: Bayesian inventory management with unobserved lost sales," *Management Science*, vol. 45, pp. 346–363, 1999.

[25] Law, A. M. and Kelton, W. D., *Simulation Modeling and Analysis*. McGraw-Hill Higher Education, 2000.

[26] Lee, H. L. and Nahmias, S., "Single product, single location-models," in *Logistics of Production and Inventory* (Graves, S., Rinnooy Kan, A., and Zipkin, P., eds.), vol. 4 of *Handbooks in Operations Research and the Management Sciences*, pp. 1–55, Amsterdam: Elsevier Science, 1993.

[27] Lee, H. L. and Ozer, O., "Unlocking the value of RFID," *Production and Operations Management*, vol. 16, pp. 40–64, 2007.

[28] Lin, Z. Z., Bean, J., and White, C. C., "A hybrid genetic/optimization algorithm for finite horizon partially observed Markov decision processes," *INFORMS Journal on Computing*, vol. 16, pp. 27–38, 2004.

[29] Lovejoy, W. S., "Computationally feasible bounds for partially observed Markov decision processes," *Operations Research*, vol. 39, pp. 162–175, 1991.

[30] Lovejoy, W. S., "A survey of algorithmic methods for partially observed Markov decision processes," *Annals of Operations Research*, vol. 28, pp. 47–66, 1991.

[31] McGillivray, A. R. and Silver, E. A., "Some concepts for inventory control under substitutable demand," *INFOR*, vol. 16, pp. 47–63, 1978.

[32] Meuleau, N., Peshkin, L., Kim, K., and Kaelbling, L. P., "Learning finite-state controllers for partially observable environments," *In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. Computer Science Dept., Brown University, Morgan Kaufmann*, 1999.

[33] Monahan, G. E., "A survey of partially observable Markov decision processes: theory, models, and algorithms," *Management Science*, vol. 28, pp. 1–16, 1982.

[34] Nagarajan, M. and Rajagopalan, S., "Inventory models for substitutable products: Monopoly and duopoly analysis," *Working Paper*, 2005.

[35] Nahmias, S., "Demand estimation in lost sales inventory systems," *Naval Research Logistics*, vol. 41, pp. 739–757, 1994.

[36] Netessine, S. and Rudi, N., "Centralized and competitive inventory models with demand substitution," *Operations Research*, vol. 51, pp. 329–335, 2003.

[37] Parlar, M. and Goyal, S., "Optimal ordering decisions for two substitutable products with stochastic demand," *OPSEARCH*, vol. 21, pp. 1–15, 1984.

[38] Parr, R. and Russell, S., "Approximating optimal policies for partially observable stochastic domains," *In Proceedings of the International Joint Conference on Artificial Intelligence, Morgan Kaufmann*, pp. 1088–1094, 1995.

[39] Pasternack, B. A. and Drezner, Z., "Optimal inventory policies for substitutable commodities with stochastic demand," vol. 38, pp. 221–240, 1991.

[40] Puterman, M. L., *Markov Decision Processes*. New York: John Wiley and Sons Inc., 1994.

[41] Rao, U. S., Swaminathan, J. M., and Zhang, J., "Mult-product inventroy planning with downward substitution, stochastic demand and setup costs," *IIE Transactions*, vol. 36, pp. 59–71, 2004.

[42] Sallans, B., "Learning factored representations for partially observable Markov decision processes.," *In Advances in Neural Information Processing Systems. The MIT Press*, vol. 12, 2000.

[43] Scarf, H., "Bayes solution to the statistical inventory problem," *Annals of Mathematical Statistics*, vol. 30, pp. 490–508, 1959.

[44] Scarf, H., "Some remarks on Bayes solutions to the inventory problem," *Naval Research Logistics Quarterly*, vol. 7, pp. 591–596, 1960.

[45] Smallwood, R. D. and Sondik, E. J., "The optimal control of partially observable Markov decision processes over a finite horizon," *Operations Research*, vol. 21, pp. 1071–1088, 1973.

[46] Smith, S. A. and Agrawal, N., "Management of multi-item inventory systems with demand substitution," *Operations Research*, vol. 48, pp. 50–64, 2000.

[47] Sondik, E. J., *Optimal Control of Partially-Observable Markov Processes*. PhD thesis, Engineering-Economic Systems, Stanford University, Stanford, California, 1971.

[48] Sondik, E. J., "The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs," *Operations Research*, vol. 2, pp. 282–304, 1978.

[49] Striebel, C. T., "Sufficient statistics in the optimal control of stochastic systems," *Journal of Mathematical Analysis and Applications*, vol. 12, pp. 576–592, 1965.

[50] Suematsu, N. and Hayashi, A., "A reinforcement learning algorithm in partially observable environments using short-term memory," *Production and Operations Management*, vol. 11, 1999.

[51] Thrun, S., "Monte Carlo POMDPs," *In Advances in Neural Information Processing Systems. The MIT Press*, vol. 12, 2000.

[52] Treharne, J. T. and Sox, C. R., "Adaptive inventory control for nonstationary demand and partial information," *Management Science*, vol. 48, pp. 607–624, 2002.

[53] Uckun, C., Karaesmen, F., and Savas, S., "Investment in improved inventory accuracy in a decentralized supply chain," *Graduate Schools of Sciences and Engineering. Koc University Working Paper*, 2005.

[54] White, C. C., "A survey of solution techniques for the partially observed Markov decision process," *Annals of Operations Research*, vol. 32, pp. 215–230, 1991.

[55] White, C. C. and Harrington, D., "Application of Jensen's inequality for adaptive suboptimal design," *Journal of Optimization Theory and Applications*, vol. 32, pp. 89–99, 1980.

[56] White, C. C. and Scherer, W., "Solution procedures for partially observed Markov decision processes," *Operations Research*, vol. 37, pp. 791–797, 1989.

[57] White, C. C. and Scherer, W., "Finite-memory suboptimal design partially observed Markov decision processes," *Operations Research*, vol. 42, pp. 439–455, 1994.

[58] Zhand, X. and Chen, J., "Joint replenishment policy for inventory system with demand substitution," *Proceedings of the 5th World Congress on Intelligent Control and Automation. June 15-19, 2004, Hangzhou P.R. China*, 2004.