

# On the Optimal Control of Partially Observed Inventory Systems \*

Alain Bensoussan<sup>†</sup>, Metin Çakanyıldırım, Suresh P. Sethi<sup>‡</sup>

International Center for Decision and Risk Analysis  
School of Management  
P.O.Box 830688, SM 30  
University of Texas at Dallas  
Richardson, TX 75083-0688

May 24, 2005

## Abstract

This paper introduces recent developments in the analysis of inventory systems with partial observations. The states of these systems are typically conditional distributions, which evolve in infinite dimensional spaces over time. Our analysis involves introducing unnormalized probabilities to transform nonlinear state transition equations to linear ones. With the linear equations, the existence of the optimal feedback policies are proved for two models where demand and inventory are partially observed. In a third model where the current inventory is not observed but a past inventory level is fully observed, a sufficient statistic is provided to serve as a state. The last model serves as an example where a partially observed model has a finite dimensional state. In that model, we also establish the optimality of the basestock and  $(s, S)$  policies, hence generalizing the corresponding classical models with full information.

## Résumé

On présente un certain nombre de modèles de systèmes de stocks avec information partielle. Ils sont formalisés comme des problèmes de contrôle où l'état est une probabilité conditionnelle, dans un espace de dimension infinie. On introduit des probabilités non normalisées, permettant de transformer des équations non linéaires en équations linéaires. On peut alors montrer l'existence de feedbacks optimaux pour deux modèles où la demande et le stock sont partiellement observables. Dans un troisième modèle, le stock n'est pas observé, mais un stock antérieur est observé. Une statistique exhaustive est obtenue, et l'état est de dimension finie. On établit l'optimalité des politiques "stock de base" et  $(s, S)$ , généralisant les modèles classiques avec information complète.

---

\*Recherche Opérationnelle/ Operations Research, présentée par Alain Bensoussan

<sup>†</sup>Professeur Emerite, Université Paris Dauphine

<sup>‡</sup>{alain.bensoussan, metin, sethi}@utdallas.edu

## Version Française Abrégée

Bien que la gestion des stocks ait été l'objet d'une littérature considérable, la situation où l'information est partielle n'a pratiquement pas été abordée jusqu'à présent. On montre, sur un certain nombre de modèles, comment l'idée très féconde des probabilités conditionnelles non normalisées, introduites dans le filtrage non linéaire par Zakai peut être adaptée et permet d'établir l'existence de politiques optimales. Les problèmes à traiter sont du type suivant

$$\pi_{t+1}(x) = \mathbb{I}_{z_t=q_t} \frac{\int_{q_t}^{\infty} \pi_t(\xi) p(x|\xi) d\xi}{\int_{q_t}^{\infty} \pi_t(\xi) d\xi} + \mathbb{I}_{z_t < q_t} p(x|z_t), \quad (1)$$

où  $p(x|\xi)$  est la probabilité de transition d'une chaîne de Markov,  $z_t$  est le processus observé,  $q_t$  le contrôle adapté à  $\mathcal{Z}_{t-1}$ .  $\pi_t(x)$  est une probabilité conditionnelle et représente l'état du système. L'équation de Bellman associée est donnée par

$$V(\pi) = \min_q \left\{ \int L(x, q) \pi(x) dx + \alpha V \left( \frac{\int_q^{\infty} p(\cdot|\xi) \pi(\xi) d\xi}{\int_q^{\infty} \pi(\xi) d\xi} \right) \int_q^{\infty} \pi(\xi) d\xi + \alpha \int_0^q V(p(\cdot|\xi)) \pi(\xi) d\xi \right\}. \quad (2)$$

La linéarisation est obtenue par les équations (8), (10), (11) ci dessous. Un principe de contraction peut s'appliquer à (8) et permet de résoudre aussi (2), et d'obtenir un feedback optimal. Un problème beaucoup plus complexe est décrit par les équations (25), (32), (33). Là aussi, la linéarisation simplifie considérablement, même si le principe de contraction ne s'applique plus. On prouve l'existence d'une solution maximale qui s'interprète comme la fonction valeur d'un problème de contrôle. On construit une suite de contrôles feedback formant une suite minimisante (le feedback optimal peut ne pas être atteint).

## 1 Introduction

Inventory control is one of the most important topics in operations research. A systematic analysis of inventory problems began with the development of the classical EOQ formula of Harris in 1913 [9] and a significant plateau was reached in 1958 by Arrow, Karlin and Scarf [1]. A critical assumption in this vast literature has been that the level of inventory at any given time is fully observed. Some of the most celebrated results such as the optimality of base-stock or  $(s, S)$  policies have been obtained under the assumption of full observation. Most of the well-known inventory policies are not only non-optimal, but are also not applicable in the partial observation environment. While one works with a finite dimensional state space in the full observation case, one usually has to deal with an infinite dimensional state space in the partial observation setting. In particular, in the full observation cases the inventory level is often the state variable, whereas with the partial observations, the system state is typically the conditional distribution of the inventory level given the partial observations.

It is the purpose of this paper to introduce some techniques, reminiscent of the Zakai equation [13] in stochastic control, for the analysis of inventory control with partial observations. The Zakai equation uses a transformation that changes the highly nonlinear Kushner equation [10] for evolution of probability distributions over time into a linear equation. While the transformation does not remove the infinite dimensionality, the linearity permits a number of important control problems with partial observations to be solved. On the other hand, some inventory models can admit a finite dimensional sufficient statistic. For a specific inventory model, this statistic turns out to be the known inventory level in an earlier period plus all of the inventory orders made

since that period. When there exists a sufficient statistic, the well-known optimal ordering policies are more likely to remain valid.

In Sections 2 and 3, we introduce two separate inventory models, both of which benefit from the derivation of a Zakai-type equation. In Section 4, we provide an inventory model which admits a sufficient statistic. Sections 2, 3 and 4 are respectively based on [5], [3] and [4].

## 2 Partially Observed Demand due to Censoring

Although many inventory models assume a static and a known demand distribution, this distribution in reality can change over time and so it is likely to be unknown. The distribution needs to be updated based on the demand observations. Ding et al. [6] study a multiperiod newsvendor model with Bayesian updating of demands. Specifically, the demand is observed *only* when it is less than the inventory level, i.e., unmet demand is *censored*. Consequently, the state of the system is the distribution of the demand characterized by a parameter, which is updated in each period based on the partial observation of the demand at that time. Ding et al. assume that the demands are independently and identically distributed. On the contrary, in this section we do not make any assumption on the demand distribution or correlation between the demands in different periods, except that it is a Markov process. Furthermore, we develop a Zakai-type equation [13] for the evolution of the conditional probability distribution of the demand over time. This facilitates the analysis of the dynamic programming equation for the problem. We establish that the value function is the unique solution of the DP equation and that there exists an optimal feedback policy for the problem.

We now formally define our problem. Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be the probability space and let  $t \geq 1$  be the indices for the periods. Let  $D_t \geq 0$  denote the demand occurring at the beginning of period  $t$ . The demand is modelled by a Markov process with the transition probabilities given by  $p(x, \xi) := \mathbb{P}(D_{t+1} = x | D_t = \xi)$ . The inventory available to satisfy the demand  $D_t$ , or a part thereof, is called  $q_t$ . Like [6], we assume that the excess inventory at the end of each period is salvaged and the unmet demand in a period is lost. Thus,  $q_t$  can also be thought as the order placed and delivered at the beginning of period  $t$  before the demand  $D_t$  arrives. Then the amount  $z_t$  of sales is given by

$$z_t := \min\{D_t, q_t\}. \quad (3)$$

When  $D_t < q_t$ , the demand is met and therefore observed. On the other hand, when  $D_t \geq q_t$ , the inventory is not sufficient to meet the demand in period  $t$ . Then, the amount of sales is only  $q_t$  and  $D_t - q_t$  is the unmet demand. When the demand is not met, the unmet demand is not observed by the inventory manager (IM). Let  $\mathcal{Z}_t$  denote the sigma algebra  $\mathcal{Z}_t := \sigma(\{z_1, \dots, z_t\})$ . Since  $\mathcal{Z}_t \subseteq \sigma(\{D_1, \dots, D_t\})$ , we have a partially observed inventory model.

Let the function  $L(D, q)$ , which depends on the demand  $D$  and the available inventory level  $q$ , denote the one-period cost function:

$$L(D, q) = \begin{cases} cq - h(q - D) & \text{if } D \leq q \\ cq + b(D - q) & \text{if } q \leq D \end{cases}, \quad (4)$$

where  $h, c$  and  $b$  are, respectively, the salvage value per unit, the ordering cost per unit, and the shortage cost per unit. It is reasonable to assume that  $0 \leq h < c < b$ .

With the discount factor  $0 < \alpha < 1$  and with  $\mathbf{q}$  defining the sequence of order quantities  $\mathbf{q} = \{q_1, q_2, \dots\}$ , with  $q_t$  adapted to  $\mathcal{Z}_{t-1}$ , our objective is to minimize

$$J(\mathbf{q}) := \sum_{t=1}^{\infty} \alpha^{t-1} \mathbb{E} L(D_t, q_t). \quad (5)$$

We take the initial distribution of  $D_1$  as given to start the process. Note that  $q_1$  is determined with certainty at the beginning of period 1. This completes the definition of our problem.

We start by studying the evolution of the demand distribution. Let  $\pi_t$  denote the conditional probability density function of  $D_t$ . Since  $\pi_{t+1}$  is updated from  $\pi_t$  and  $\mathcal{Z}_t$ , it is obvious that  $\pi_{t+1}$  is  $\mathcal{Z}_t$ -measurable. When an integral is taken over  $[0, \infty)$ , we suppress the limits to save on the notation.

Using probabilistic arguments, we obtain the random density measure  $\pi_{t+1}$  in terms of  $\pi_t$ :

$$\pi_{t+1}(x) = \mathbb{1}_{z_t=q_t} \frac{\int_{q_t}^{\infty} \pi_t(\xi) p(x|\xi) d\xi}{\int_{q_t}^{\infty} \pi_t(\xi) d\xi} + \mathbb{1}_{z_t < q_t} p(x|z_t). \quad (6)$$

With this distribution as the state variable, we now proceed to derive the dynamic programming equation for our problem.

## 2.1 Bellman equation with normalized and unnormalized probabilities

We begin with the expected total costs occurring in the current period  $t$ , i.e.,  $\mathbb{E} L(D_t, q_t) = \int L(x, q_t) \pi_t(x) dx$ . Equation (6) implies that the future costs should have two terms depending on whether the demand is observed. When the demand is not observed, the future costs depend on the first term on the right-hand side of (6). Otherwise, they depend on the second term. Summing up the current and future costs that result in by ordering  $q$  yields the terms inside the curly brackets in the Bellman equation below:

$$V(\pi) = \min_q \left\{ \int L(x, q) \pi(x) dx + \alpha V \left( \frac{\int_q^{\infty} p(\cdot|\xi) \pi(\xi) d\xi}{\int_q^{\infty} \pi(\xi) d\xi} \right) \int_q^{\infty} \pi(\xi) d\xi + \alpha \int_0^q V(p(\cdot|\xi)) \pi(\xi) d\xi \right\}. \quad (7)$$

This equation yields the optimal  $q$  as a function of the current knowledge summarized in  $\pi$  about the demand. The state transition equations can be linearized as we illustrate next.

We begin with (6) and develop what is known as the unnormalized probability. Its dynamics is similar to the Zakai equation obtained in the filtering literature [13].

Taking a cue from (6), we define  $\rho_t(x)$  by the recursive linear equation

$$\rho_{t+1}(x) = \mathbb{1}_{z_t=q_t} \int_{q_t}^{\infty} p(x|\xi) \rho_t(\xi) d\xi + \mathbb{1}_{z_t < q_t} p(x|z_t) \quad \text{for } t \geq 1, \quad (8)$$

with the initial value  $\rho_1(x) = \rho(x) = \pi(x)$ . Equation (8) corresponds to the Zakai equation for systems with diffusions in [13] and [2]. Also set  $\lambda_t := \int \rho_t(x) dx$ . Then we have

$$\lambda_1 = 1, \quad \lambda_{t+1} = \mathbb{1}_{z_t=q_t} \int_{q_t}^{\infty} \rho_t(\xi) d\xi + \mathbb{1}_{z_t < q_t} \quad \text{for } t \geq 1, \quad (9)$$

where (9) follows directly from integrating the left-hand side of (8) over  $[0, \infty)$ . Moreover,  $\rho_t(x) = \pi_t(x) \lambda_t$ , where  $\lambda_t$  is a weighting factor which allows us to recover normalized probabilities.

We define, for any positive  $L^1$  function  $\rho$ ,

$$W(\rho) := V\left(\frac{\rho}{\int \rho(x)dx}\right) \int \rho(x)dx. \quad (10)$$

We use (7) and (10) to obtain a recursive equation for  $W$ :

$$W(\rho) = \min_q \left\{ \int L(x, y)\rho(x)dx + \alpha W\left(\int_q^\infty p(\cdot|\xi)\rho(\xi)d\xi\right) + \alpha \int_0^q W(p(\cdot|\xi))\rho(\xi)d\xi \right\}. \quad (11)$$

From the linearity in  $\rho$  of the first term inside the minimum above, we can see that  $W(0) = 0$  and  $W$  is homogenous of degree 1, i.e.,  $W(a\rho) = aW(\rho)$  for  $a > 0$ . The right-hand side in the Bellman equation (11) has linear growth, which facilitates the derivation of the existence results below.

## 2.2 Existence of an Optimal Feedback Solution and Monotone Value Functions

First we show that there exists a unique solution of the DP equation. For this, we define some normed linear spaces. Let  $\mathcal{H} := \{\rho \in L^1(\mathfrak{R}^+) : \int x|\rho(x)|dx < \infty\}$ , where  $L^1(\mathfrak{R}^+)$  is the space of integrable functions whose domain is the set of nonnegative real numbers.  $\mathcal{H}$  is a subspace of  $L^1(\mathfrak{R}^+)$ . Since we are interested in nonnegative unnormalized probabilities, we define  $\mathcal{H}^+ = \{\rho \in \mathcal{H} | \rho \geq 0\}$ , a closed subset of  $\mathcal{H}$  with the norm  $\|\rho\| = \int |\rho(x)|dx + \int x|\rho(x)|dx$ . Let us also define the space  $\mathcal{B} = \{\phi(\rho) : \mathcal{H}^+ \rightarrow \mathfrak{R} \mid \sup_{x>0} |\phi(\rho)|/\|\rho\| < \infty\}$  with the norm  $\|\phi\|_{\mathcal{B}} = \sup_{\rho \in \mathcal{H}^+} |\phi(\rho)|/\|\rho\|$ .

We assume that for any  $\mu \in \mathcal{B}$ ,  $\int x \int p(x|\xi)\mu(\xi)d\xi dx \leq \bar{c} \int \xi \mu(\xi)d\xi$  with  $\alpha \bar{c} < 1$ . We also define the map  $T(W)$  as

$$T(W)(\rho) := \min_q \left\{ \int L(x, y)\rho(x)dx + \alpha W\left(\int_q^\infty p(\cdot|\xi)\rho(\xi)d\xi\right) + \alpha \int_0^q W(p(\cdot|\xi))\rho(\xi)d\xi \right\}. \quad (12)$$

The next theorem summarizes the results.

**Theorem 1.** *If there exists a solution  $W(\rho)$  of (11), it is in  $\mathcal{B}$ . Furthermore,  $\|T(W) - T(\tilde{W})\|_{\mathcal{B}} \leq \alpha \max\{1, \bar{c}\} \|W - \tilde{W}\|_{\mathcal{B}}$ . Thus, there exists one and only one solution  $W(\rho)$  of the Bellman equation (11). Moreover,  $W(\rho)$  is continuous at each  $\rho \in \mathcal{H}^+$  and there exists an optimal feedback policy.*

Using a value iteration based on (11), we can deduce that  $W(\rho)$  is monotone increasing and concave in  $\rho$ . More interestingly, the total discounted cost  $V(\pi)$  is smaller when the demand is believed to be smaller in the hazard rate order; see e.g., Definition 1.3.2 of [11] for this order. We also say that linear operator  $\mathcal{P}(\pi)$  defined by  $\mathcal{P}(\pi) := \int p(\cdot|\xi)\pi(\xi)d\xi$  is hazard rate order preserving if  $\mathcal{P}(\pi) \preceq \mathcal{P}(\pi')$  holds for all  $\pi$  and  $\pi'$  with  $\pi \preceq \pi'$ .

**Theorem 2.**  *$V(\pi) \leq V(\pi')$  if  $\pi$  is smaller than  $\pi'$  in the hazard rate order and  $\mathcal{P}(\pi)$  is hazard rate order preserving.*

## 2.3 Independently and identically distributed demands

Here we consider the case of i.i.d. demands parameterized by  $\tau$ . The parameter  $\tau$  is unknown, and we use the Bayesian approach to learn about  $\tau$ . Let  $f(\cdot|\tau)$  be the probability density function of the demand  $D$ , while the process  $\{D_t : t \geq 1\}$  is made up of independent copies of  $D$ . The observation process, the one-period cost and the objective still are defined respectively by (3), (4) and (5).

Let  $\varpi_t$  be the distribution of the parameter  $\tau$  before observing  $z_t$ . That is,  $\varpi_t$  is a prior used to build the distribution of  $D_t$ :  $P(D_t = x|\varpi_t) = \int f(x|\tau)\varpi_t(\tau)d\tau$ . We start with a known  $\varpi_1$  and obtain the evolution of the process  $\{\varpi_t : t \geq 1\}$ :

$$\varpi_{t+1}(\tau) = \mathbb{1}_{z_t=q_t} \frac{\int_{q_t}^{\infty} f(\xi|\tau)\varpi_t(\tau)d\xi}{\int_{q_t}^{\infty} f(\xi|\tau)\varpi_t(\tau)d\xi d\tau} + \mathbb{1}_{z_t < q_t} \frac{f(z_t|\tau)\varpi_t(\tau)}{\int f(z_t|\tau)\varpi_t(\tau)d\tau}. \quad (13)$$

This evolution equation is similar to (6).

Using the distribution of  $\tau$ , we easily obtain the distribution  $\pi_t$  of demand  $D_t$  as  $\int f(x|\tau)\varpi_t(\tau)d\tau$ . This distribution helps us to compute the current period costs while the future costs are obtained by using (13), so the Bellman equation becomes

$$V(\varpi) = \min_q \left\{ \int \int L(x, q) f(x|\tau)\varpi(\tau)d\tau dx + \alpha V \left( \frac{\int_q^{\infty} f(\xi|\cdot)\varpi(\cdot)d\xi}{\int \int_q^{\infty} f(\xi|\tau)\varpi(\tau)d\xi d\tau} \right) \int_q^{\infty} \int f(x|\tau)\varpi(\tau)d\tau dx \right. \\ \left. + \alpha \int_0^q V \left( \frac{f(x|\cdot)\varpi(\cdot)}{\int f(x|u)\varpi(u)du} \right) \int f(x|\tau)\varpi(\tau)d\tau dx \right\}. \quad (14)$$

Similar to (8), the unnormalized probability  $\rho_t$  for  $\tau$  evolves as follows:

$$\rho_{t+1}(\tau) = \mathbb{1}_{z_t=q_t} \int_{q_t}^{\infty} f(u|\tau)du\rho_t(\tau) + \mathbb{1}_{z_t < q_t} f(z_t|\tau)\rho_t(\tau), \quad (15)$$

with  $\rho_1 = \varpi_1$ . The Bellman equation is

$$W(\rho) = \min_q \left\{ \int \int L(x, q) f(x|\tau)\rho(\tau)d\tau dx + \alpha W \left( \int_q^{\infty} f(\xi|\cdot)\rho(\cdot)d\xi \right) + \alpha \int_0^q W(f(x|\cdot)\rho(\cdot)) dx \right\}. \quad (16)$$

Writing the Bellman equation with unnormalized probabilities facilitates obtaining results similar to those listed in Theorem 1.

### 3 Partially Observed Inventory due to Zero-Balance Walk

In this section, the demand has a known distribution but the inventory level  $I_t$  at the beginning of period  $t$  is given by

$$I_{t+1} = (I_t + q_t - D_t)^+ \quad \text{for } t \geq 1. \quad (17)$$

In every period  $t$ , demand  $D_t$  is met, to the extent possible, from the on-hand stock  $I_t + q_t$ . When the demand is met entirely, inventory holding costs apply to the remaining inventory. Otherwise, there are lost sales costs. We assume demand  $D_t$  to be i.i.d. A generic demand is denoted by  $D$ , which is i.i.d. with each  $D_t$ . Let  $f$  denote the density and  $F$  denote the cumulative distribution of  $D$ . Let  $\bar{F} = 1 - F$ .

In classical inventory settings, the inventory level  $I_t$  is observed, and is used to determine the order quantity  $q_t$  in period  $t$ . In this section, the IM does not automatically observe the inventory level due to transaction errors, misplaced inventories, spoilage or production yield. The inventory level is observed only when there is no physical inventory. When there is inventory, only the event that the inventory is positive is observed. We continue to use  $z_t$  to denote the observed signal, which in this context is

$$z_t := \mathbb{1}_{I_t=0}, \quad t \geq 0. \quad (18)$$

The signal  $z_t$  is a discrete-time Markov Chain with the state space  $\{0, 1\}$ : 1 means an empty inventory shelf and 0 means a nonempty shelf. This observation process mimics what is known as “zero-balance walk” ([8] and [12]) at some companies where employees walk around the shelves to identify the stocked-out items and verify the inventory levels for those items.

The order  $q_t$  is adapted to  $\mathcal{Z}_t := \sigma(\{z_j : 1 \leq j \leq t\})$ . Clearly  $\mathcal{Z}_t \subset \mathcal{F}_t := \sigma(\{I_j : 1 \leq j \leq t\})$ . Given a stationary cost function  $c(I_t, q_t)$  that depends on the inventory level  $I_t$  and the order size  $q_t$ , and with  $\mathbf{q}$  defining the admissible order quantities, the total discounted cost is defined by

$$J(\zeta, \pi, \mathbf{q}) := \sum_{t=1}^{\infty} \alpha^{t-1} \mathbb{E} c(I_t, q_t). \quad (19)$$

The initial conditions are a pair  $(\zeta, \pi(x))$ , where  $\zeta$  is 1 or 0. If  $\zeta$  is 1, then  $I_1 = 0$ . If  $\zeta$  is 0, then  $I_1 > 0$  and  $\pi(\cdot)$  is the probability distribution of  $I_1$ . We look for  $q_t$ , adapted to  $\mathcal{Z}_t$ ,  $t \geq 0$ , to minimize  $J(\zeta, \pi, \mathbf{q})$ .

### 3.1 Evolution of State Probabilities

We now develop the conditional probability density  $\pi_t(\cdot)$  of  $I_t$  given  $\mathcal{Z}_{t-1}$  and  $I_t > 0$ . Derivation of  $\pi_t$  in this case is more involved than in Section 2, so we furnish more details. By definition,  $\int_0^x \pi_t(y) dy = P(I_t \leq x | \mathcal{Z}_{t-1}, I_t > 0)$ . Since the event  $[I_t = 0]$  is observable, conditional probabilities are needed only when  $I_t > 0$ . We can use the conditional Bayes theorem (e.g. [7]) to obtain

$$\int_0^{\infty} \phi(x) \pi_t(x) dx = \mathbb{E}[\phi(I_t) | \mathcal{Z}_{t-1}, I_t > 0] = \frac{\mathbb{E}[\phi(I_t) \mathbb{I}_{I_t > 0} | \mathcal{Z}_{t-1}]}{\mathbb{E}[\mathbb{I}_{I_t > 0} | \mathcal{Z}_{t-1}]}, \quad (20)$$

where  $\phi$  is an arbitrary test function. In order to obtain a recursive expression for  $\pi_t$  in terms of  $\pi_{t-1}$ , we begin with expressing  $\mathbb{E}(\phi(I_t) | \mathcal{Z}_t)$  in terms of conditional expectations with respect to  $\mathcal{Z}_{t-1}$  in the next two lemmas.

**Lemma 1.**

$$\mathbb{E}(\phi(I_t) | \mathcal{Z}_t) = \mathbb{I}_{I_t=0} \phi(0) + \mathbb{I}_{I_t>0} \frac{\mathbb{E}(\phi(I_t) \mathbb{I}_{I_t>0} | \mathcal{Z}_{t-1})}{P(I_t > 0 | \mathcal{Z}_{t-1})} = \mathbb{I}_{I_t=0} \phi(0) + \mathbb{I}_{I_t>0} \mathbb{E}(\phi(I_t) | \mathcal{Z}_{t-1}, I_t > 0). \quad (21)$$

**Lemma 2.**

$$\begin{aligned} \mathbb{E}(\phi(I_t) | \mathcal{Z}_t) \mathbb{I}_{I_t>0} &= \mathbb{I}_{I_{t-1}=0} \frac{\int_0^{\infty} \phi(z) f(q_{t-1} - z) \mathbb{I}_{q_{t-1} \geq z} dz}{F(q_{t-1})} \\ &+ \mathbb{I}_{I_{t-1}>0} \frac{\int_0^{\infty} \phi(z) \int_{(z-q_{t-1})^+}^{\infty} f(y + q_{t-1} - z) \pi_{t-1}(y) dy dz}{\int_0^{\infty} F(y + q_{t-1}) \pi_{t-1}(y) dy}. \end{aligned} \quad (22)$$

Going back to the conditional probability  $\pi_t$ , we may write

$$\mathbb{E}(\phi(I_t) | \mathcal{Z}_t) = \mathbb{I}_{I_t=0} \phi(0) + \mathbb{I}_{I_t>0} \int_0^{\infty} \phi(z) \pi_t(z) dz. \quad (23)$$

Setting the second term on the right-hand side of (23) equal to (22), we get

$$\pi_t(x) = \mathbb{I}_{I_{t-1}=0} \left\{ \frac{f(q_{t-1} - x) \mathbb{I}_{x \leq q_{t-1}}}{F(q_{t-1})} \right\} + \mathbb{I}_{I_{t-1}>0} \left\{ \frac{\int_{(x-q_{t-1})^+}^{\infty} f(y + q_{t-1} - x) \pi_{t-1}(y) dy}{\int_0^{\infty} F(y + q_{t-1}) \pi_{t-1}(y) dy} \right\}. \quad (24)$$

Equivalently, the conditional probability evolves according to the highly nonlinear equation

$$\pi_t(x) = z_{t-1} \frac{f(q_{t-1} - x) \mathbb{1}_{x < q_{t-1}}}{F(q_{t-1})} + (1 - z_{t-1}) \frac{\int_{(x - q_{t-1})^+}^{\infty} f(y + q_{t-1} - x) \pi_{t-1}(y) dy}{\int_0^{\infty} F(q_{t-1} + y) \pi_{t-1}(y) dy}, \quad t \geq 2, \quad \pi_1(x) = \pi(x), \quad (25)$$

which corresponds to the Kushner equation [10] in our inventory context.

For linearizing (25), we set

$$\rho_t(x) := \lambda_t \pi_t(x). \quad (26)$$

On account of this weighting,  $\rho_t(x)$  becomes the unnormalized probability and it evolves according to the linear equation

$$\rho_t(x) = z_{t-1} f(q_{t-1} - x) \mathbb{1}_{x < q_{t-1}} + (1 - z_{t-1}) \int_{(x - q_{t-1})^+}^{\infty} f(y + q_{t-1} - x) \rho_{t-1}(y) dy, \quad \rho_1(x) = \pi(x). \quad (27)$$

By integrating both sides of (26), we get

$$\lambda_t = z_{t-1} F(q_{t-1}) + (1 - z_{t-1}) \lambda_{t-1} \int F(q_{t-1} + y) \pi_{t-1}(y) dy,$$

also a linear equation defining  $\lambda_t$  recursively starting with  $\lambda_1 = 1$ . However, note that  $\lambda_t$  depends on  $\pi_{t-1}$  on the right-hand side. The normalized probabilities can easily be computed from the unnormalized probabilities by  $\pi_t(x) = \rho_t(x) / \int \rho_t(x) dx$ .

The evolution equations can be written in the operator form in the space  $\mathcal{H}$  and its dual space  $\mathcal{H}_* = \{\phi : \sup_{x > 0} |\phi(x)| / (1 + x) < \infty\}$ . Furthermore,  $\langle \rho, \phi \rangle = \int_0^{\infty} \rho(x) \phi(x) dx$  for  $\rho \in \mathcal{H}$ ,  $\phi \in \mathcal{H}_*$ .

For any scalar  $q > 0$ , define the linear operator  $\Phi$  from  $\mathcal{H}$  to  $\mathcal{H}$  as

$$\Phi(q, \rho)(x) = \int_{(x - q)^+}^{\infty} f(y + q - x) \rho(y) dy. \quad (28)$$

Note that  $\Phi(q, \delta)(x) = f(q - x) \mathbb{1}_{x < q}$  so  $\Phi(0, \delta)(x) = 0$  for the Dirac delta function  $\delta$ . Define the nonlinear operator  $\Psi$  as

$$\Psi(q, \rho) = \frac{\Phi(q, \rho)}{\langle \Phi(q, \rho), 1 \rangle}. \quad (29)$$

With these notations, we can write (25) and (27) in the operator form:

$$\pi_t = z_{t-1} \Psi(q_t, \delta) + (1 - z_{t-1}) \Psi(q_t, \pi_{t-1}), \quad (30)$$

$$\rho_t = z_{t-1} \Phi(q_t, \delta) + (1 - z_{t-1}) \Phi(q_t, \rho_{t-1}), \quad (31)$$

with the initial conditions  $\pi_1 = \rho_1 = \pi$ . Once again, we emphasize that (31) is a linear equation, while (30) is nonlinear.

### 3.2 The Bellman Equation

We write  $\pi_t[\mathbf{q}]$  to emphasize the dependence of the state  $\pi_t$  on the control policy. We assume that  $c(I_t, q_t)$  has linear growth in  $I_t$  for every fixed  $q_t$ , i.e.,  $c(\cdot, q_t) \in \mathcal{H}_*$ . The cost function is:

$$J(\zeta, \pi, \mathbf{q}) = \sum_{t=1}^{\infty} \alpha^{t-1} \mathbb{E}[\mathbb{E}[c(I_t, q_t) | \mathcal{Z}_t]] = \sum_{t=1}^{\infty} \alpha^{t-1} \mathbb{E}\{z_t c(0, q_t) + (1 - z_t) \langle c(I_t, q_t), \pi_t[\mathbf{q}] \rangle\}$$



where  $\pi_t[\mathbf{q}]$  is given by (25). Recall that the initial conditions  $\zeta_1 = \zeta \in \{0, 1\}$  and  $\pi_1 = \pi$  are given. We now study only the discounted infinite horizon cost, so the time index  $t$  is suppressed. We define the value function

$$V(\zeta, \pi) := \inf_{\mathbf{q}} J(\zeta, \pi, \mathbf{q}).$$

If we write  $v := V(1, \pi)$  which, in fact, is not dependent on  $\pi$ , and  $V(\pi) := V(0, \pi)$ , then we obtain the following system:

$$V(\pi) = \inf_q \left\{ \langle c(\cdot, q), \pi(\cdot) \rangle + \alpha v \int \bar{F}(y+q) \pi(y) dy + \alpha V(\Psi(q, \pi)) \int F(y+q) \pi(y) dy \right\}, \quad (32)$$

$$v = \inf_q \left\{ c(0, q) + \alpha v \bar{F}(q) + \alpha V(\Psi(q, \delta)) \int F(q) \right\}. \quad (33)$$

The study of the system in (32)-(33) simplifies considerably when working with the unnormalized probability  $\rho \in \mathcal{H}^+$ . Towards that end, we define a new value function  $Z(\cdot)$ :  $Z(\rho) := V(\rho/\lambda) \lambda$  and  $\lambda := \int \rho(x) dx$ . We obtain the following new system of equations:

$$Z(\rho) = \inf_q \left\{ \langle c(\cdot, q), \rho(\cdot) \rangle + \alpha v \int \bar{F}(y+q) \rho(y) dy + \alpha Z(\Phi(q, \rho)) \right\}, \quad (34)$$

$$v = \inf_q \left\{ c(0, q) + \alpha v \bar{F}(q) + \alpha Z(\Phi(q, \delta)) \right\}. \quad (35)$$

The pair  $(v, Z(\rho))$  is the solution of (34)-(35). We have  $Z(0) = 0$  and  $Z$  is homogenous of degree 1. Unlike the operator  $\Psi$ ,  $\Phi$  is a linear operator.

### 3.3 Existence of a Solution to the Bellman Equation

For the existence results, we bound the single period cost. Suppose that positive constants  $c, c_0, c_1$ , and  $h$  are such that  $cq < c(x, q) \leq c_0 + c_1q + hx$  for  $x \geq 0$ , where  $c_0$  can be interpreted as the maximum expected lost sales cost that can be incurred in a period. Indeed, we set  $c_0 = c(0, 0)$ .

We need some short-hand notation: Define the function  $K : \mathfrak{R} \times \mathcal{H} \rightarrow \mathfrak{R}$  as  $K(q, \rho; v, Z) := \langle c(\cdot, q), \rho(\cdot) \rangle + \alpha v \int \bar{F}(y+q) \rho(y) dy + \alpha Z(\Phi(q, \rho))$ . For  $\rho = \delta$ ,  $K(q, \delta; v, Z) = c(0, q) + \alpha v \bar{F}(q) + \alpha Z(\Phi(q, \delta))$ . Define the map  $T : \mathfrak{R} \times \mathcal{B} \rightarrow \mathfrak{R} \times \mathcal{B}$  as  $T(v; Z(\rho)) := (\inf_q K(q, \delta; v, Z); \inf_q K(q, \rho; v, Z))$ . Define  $(Z_0(\rho), v_0)$  as the value function when  $q = 0$ . Then, we have respectively

$$\langle c(\cdot, 0), \rho(\cdot) \rangle + \alpha v_0 \int \bar{F}(y) \rho(y) dy + \alpha Z_0(\rho(0, \rho)) = Z_0(\rho) \quad \text{and} \quad v_0 = c_0 + \alpha v_0. \quad (36)$$

**Lemma 3.**  $Z_0$  exists and is uniquely defined in  $\mathcal{B}$ .

Note that  $T(v_0; Z_0(\rho)) \leq (v_0; Z_0(\rho))$  and  $T(v; Z(\rho)) \leq (v; Z(\rho))$ . These inspire a value iteration scheme which yields  $(\bar{v}, \bar{Z})$ .

**Theorem 3.**  $(\bar{v}, \bar{Z})$  is a solution of (34)-(35) and  $0 \leq \bar{v} \leq v_0$ ,  $0 \leq \bar{Z}(\rho) \leq Z_0(\rho)$ . Moreover, it is the maximal solution satisfying  $(v; Z) = T(v; Z)$ . Also  $\bar{Z}(\pi) = \inf_{\mathbf{q}} J(0, \pi, \mathbf{q})$  and  $\bar{v} = \inf_{\mathbf{q}} J(1, \pi, \mathbf{q})$ .

Owing to the last statement,  $\bar{Z}(\pi)$  and  $\bar{v}$  are interpreted as the infima of the costs. This, however, does not imply the existence of a feedback policy unless the order quantity  $q$  is bounded. When the optimal  $q$  is bounded in terms of  $\rho$ , the bound can be arbitrarily large as  $\rho \rightarrow 0$ . Because of this, we choose to assume a

bound on  $q$  in addition to the cost bounds in the previous section. Let the maximum production capacity be  $m$  and let the corresponding  $Z$  and  $v$  be denoted by  $Z^m$  and  $v^m$ . Then (34)-(35) is written as

$$\begin{aligned} Z^m(\rho) &= \inf_{q \leq m} \left\{ \langle c(\cdot, q), \rho(\cdot) \rangle + \alpha v^m \int \bar{F}(y+q) \rho(y) dy + \alpha Z^m(\Phi(q, \rho)) \right\} \\ v^m &= \inf_{q \leq m} \{ c(0, q) + \alpha v^m \bar{F}(q) + \alpha Z^m(\Phi(q, \delta)) \}. \end{aligned} \quad (37)$$

We can check that constants  $A^m$  and  $B^m$  exist such that

$$|Z^m(\rho) - Z^m(\rho')| \leq A^m \int |\rho(y) - \rho'(y)| dy + B^m \int y |\rho(y) - \rho'(y)| dy \quad (38)$$

for any two  $\rho, \rho' \in \mathcal{H}$ . Therefore,  $Z^m$  is Lipschitz continuous on  $\mathcal{H}$ . This additional smoothness property allows us to establish the uniqueness of a solution to the system in (37) in the absence of a contraction property on  $T$  as in Theorem 1. The next result validates the monotone iterative process. That is,  $(v^m, Z^m)$  minimizes the total discounted cost.

**Proposition 1.** *The solution  $(v^m, Z^m)$  of (37) is the minimum total discounted cost, i.e.,  $Z^m(\pi) = \inf_{\mathbf{q}: q_t \leq m} J(0, \pi, \mathbf{q})$  and  $v^m = \inf_{\mathbf{q}: q_t \leq m} J(1, \delta, \mathbf{q})$ .*

Since  $Z^m(\pi)$  and  $v^m$  are defined as a solution of (37) and they are given by the infima in Proposition 1, both  $Z^m(\pi)$  and  $v^m$  are unique. As  $m$  increases, we have  $\inf_{\mathbf{q}: q_t \leq m} J(0, \pi, \mathbf{q}) \downarrow \inf_{\mathbf{q}} J(0, \pi, \mathbf{q})$  and  $\inf_{\mathbf{q}: q_t \leq m} J(1, \delta, \mathbf{q}) \downarrow \inf_{\mathbf{q}} J(1, \delta, \mathbf{q})$ . These convergencies imply the last statement in Theorem 3.

## 4 Partially Observed Inventory due to Information Delays

Here we consider partial observability arising from an information delay. The current inventory level is not observed by the IM. Instead, he observes the exact inventory level a random number of periods ago. Such a model enables us to find a sufficient statistic. It is important to allow for the backordering of the immediately unmet demand to obtain the sufficient statistic. Hence,  $I_{t+1} = I_t + q_t - D_t$  for  $t \geq 1$ . We suppose that the demands are i.i.d. and use  $D^i$  to denote the sum of  $i$  i.i.d. demands. Since we must initialize the inventory to study its evolution, we assume that the initial inventory  $I_1$  is known at  $t = 1$ .

Let  $\theta \geq 0$  be the amount of random delay. It is convenient to introduce a process  $\beta_t$ , which denotes the calendar time of the most recent inventory level observed in period  $t$ ,  $1 \leq \beta_t \leq t$ . The stochastic process  $\{\beta_t : t \geq 1\}$  is derived from the random delay  $\theta$  in the following way:

$$\beta_t = \max\{t - \theta, 1\} = \begin{cases} 1 & \text{if } t \leq \theta + 1 \\ t - \theta & \text{if } t \geq \theta + 2 \end{cases}, \quad \beta_1 = 1, \quad (39)$$

so that  $\beta_t \leq \beta_{t+1} \leq t + 1$ . The observed signal  $z_t$  at time  $t$  denotes the last inventory level observed in period  $t$ . Thus,  $z_t = I_{\beta_t}$ , which along with (39) imply  $z_1 = I_1$ . The signal process can be written as

$$z_t = \mathbb{1}_{t \geq \theta + 2} I_{t - \theta} + \mathbb{1}_{t \leq \theta + 1} I_1. \quad (40)$$

Thus,  $z_t = I_1$  if  $t \leq \theta + 1$ , since  $I_1$  is the last inventory level observed in this case. On the other hand,  $z_t = I_{t - \theta}$  if  $t \geq \theta + 2$ . The signal observed at the beginning of period  $t$  is the pair  $(z_t, \beta_t)$ , which generates the sigma algebra

$$\mathcal{Z}_t = \sigma(\{(z_1); (z_2, \beta_2); \dots; (z_t, \beta_t)\}) = \sigma\{(I_1); (I_{\beta_2}, \beta_2); \dots; (I_{\beta_t}, \beta_t)\}. \quad (41)$$

On the other hand, the underlying state at time  $t$  is  $(I_t, \mathbb{1}_{\theta=t-2})$ , and this process generates the sigma algebra  $\mathcal{F}_t$ . Clearly,  $\mathcal{Z}_t \subseteq \mathcal{F}_t$ .

With a random delay of  $\theta$ , the expression for the total discounted cost is  $J(\mathbf{q}) := \sum_{t=1}^T \alpha^{t-1} \mathbb{E} c(I_t, q_t)$ . The objective is to minimize  $J(\mathbf{q})$  over the class of admissible orders adapted to  $\mathcal{Z}_t$  defined in (41).

Now we define the *reference inventory position* as

$$x_t := z_t + \sum_{j=\beta_t}^{t-1} q_j. \quad (42)$$

Note that  $x_t = z_t$  when  $\beta_t = t$ . Since  $\{q_j : \beta_t \leq j \leq t-1\}$  are all measurable with respect to  $\mathcal{Z}_t$ ,  $x_t$  can be obtained from the information in  $\mathcal{Z}_t$ . The evolution equations for  $x_t$  and  $\beta_t$  are

$$x_{t+1} = x_t + q_t - \mathbb{1}_{\theta \leq t-1} D_{t-\theta} \quad \text{and} \quad \beta_{t+1} = \beta_t + \mathbb{1}_{\theta \leq t-1}. \quad (43)$$

Moreover, the unobserved  $I_t$  can be expressed through the observed  $x_t$  as follows:

$$I_t = x_t - \sum_{j=\beta_t}^{t-1} D_j. \quad (44)$$

By a sequence of conditioning arguments, we obtain

$$\mathbb{E} c(I_t, q_t) = \mathbb{E} \{ \mathbb{E}[c(I_t, q_t) | \mathcal{Z}_t] \} = \mathbb{E} \left\{ \mathbb{E} \left[ c \left( x_t - \sum_{j=\beta_t}^{t-1} D_j, q_t \right) \middle| \mathcal{Z}_t \right] \right\} = \mathbb{E} \left[ c^{t-\beta_t}(x_t, q_t) \right], \quad (45)$$

where  $c^i(x, q) := \mathbb{E} c(x - D^i, q)$ . Thus, the single-period costs can be computed in terms of the reference inventory position. By using (45), the objective can be rewritten as

$$J(\mathbf{q}) = \sum_{t=1}^T \alpha^{t-1} \mathbb{E} c(I_t, q_t) = \sum_{t=1}^T \alpha^{t-1} \mathbb{E} c^{t-\beta_t}(x_t, q_t). \quad (46)$$

This shows that  $(x_t, \beta_t)$  is a sufficient statistic for our problem. Moreover, we have a problem in the form of a standard inventory problem with the state  $(x_t, \beta_t)$ , with the state evolution dynamics (43), with the single-period nonstationary cost function  $c^{t-\beta_t}(x_t, q_t)$ , and with the information set  $\mathcal{Z}_t$ . Therefore, we can directly write the associated DP for the value functions  $V_t(x, \beta)$  as

$$\begin{aligned} V_{T+1}(x, \beta) &= \mathbb{E} r(x - D^{T-\beta}) \\ V_t(x, \beta) &= \inf_{q \geq 0} \left\{ c^{t-\beta}(x, q) + \alpha \mathbb{E} \left[ V_{t+1}(x + q - D^{\beta_{t+1}-\beta}, \beta_{t+1}) \middle| \beta_t = \beta \right] \right\} \end{aligned} \quad (47)$$

The DP in (47) can be made more explicit by examining  $\mathbb{P}(\beta_{t+1} = j | \beta_t = i)$ . Once  $\theta$  is observed, i.e.,  $\beta_t \geq 2$ ,  $\theta$  remains constant. Namely, since  $\beta_t = i \geq 2$  implies  $\beta_{t+1} = i + 1$ , we have

$$\mathbb{P}(\beta_{t+1} = j | \beta_t = i) = \begin{cases} 1 & \text{if } j = i + 1 \\ 0 & \text{if } j \neq i + 1 \end{cases} \quad \text{if } i \geq 2.$$

If  $\theta$  is not observed by period  $t$ , then  $\beta_t = 1$ . There is a chance that  $\theta$  will be observed in period  $t + 1$ . But when  $\theta$  is observed in period  $t + 1$ , it can be observed only as  $t - 1$ . In that case,  $\beta_{t+1} = t + 1 - (t - 1) = 2$ . This argument establishes that  $\mathbb{P}(\beta_{t+1} \geq 3 | \beta_t = 1) = 0$ . Let  $p_k := \mathbb{P}(\theta = k)$ . Then,

$$\mathbb{P}(\beta_{t+1} = 1 | \beta_t = 1) = \mathbb{P}(\theta \geq t | \theta \geq t - 1) = \frac{1 - \sum_{k=0}^{t-1} p_k}{1 - \sum_{k=0}^{t-2} p_k}.$$

Similarly,

$$P(\beta_{t+1} = 2 | \beta_t = 1) = \frac{p_{t-1}}{1 - \sum_{k=0}^{t-2} p_k} =: \Lambda_{t-1}.$$

Note that  $\Lambda_t$  is known as the hazard rate function in the reliability theory for a discrete-valued random variable  $\theta$ . Now the DP can be written as

$$\begin{aligned} V_t(x, \beta) = \inf_{q \geq 0} & \left\{ c^{t-\beta}(x, q) + \mathbb{I}_{\beta \geq 2} \alpha \mathbb{E} V_{t+1}(x + q - D, \beta + 1) \right. \\ & \left. + \mathbb{I}_{\beta=1} \alpha \{ \Lambda_{t-1} \mathbb{E} V_{t+1}(x + q - D, 2) + (1 - \Lambda_{t-1}) V_{t+1}(x + q, 1) \} \right\}. \end{aligned} \quad (48)$$

where  $\beta \leq t$ .

The one-period cost  $c^i(x, q)$  can typically be written as  $cq + \mathbb{E} H(x - D^i)$ , where the first term  $c \geq 0$  (second term  $H(\cdot)$ ) represents the ordering (inventory) cost. Suppose that the inventory costs are charged at the beginning of a period and let

$$u_t^*(\beta) := \arg \min_u [cw + \mathbb{I}_{\beta \geq 2} \alpha \mathbb{E} V_{t+1}(u - D, \beta + 1) + \mathbb{I}_{\beta=1} \alpha \{ \Lambda_{t-1} \mathbb{E} V_{t+1}(u - D, 2) + (1 - \Lambda_{t-1}) V_{t+1}(u, 1) \}].$$

With this cost structure, we have the next theorem.

**Theorem 4.** *The optimal ordering policy is state  $\beta$  dependent base stock policy, i.e., the optimal order quantity is  $\max\{u_t^*(\beta) - x, 0\}$ .*

Theorem 4 can be extended to show the optimality of an  $(s, S)$  policy, if a fixed ordering cost is charged in addition to  $c(I, q)$ .

## References

- [1] K.J. Arrow, S. Karlin and H.E. Scarf (1958). *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, California.
- [2] A. Bensoussan (1992). *Stochastic control of partially observable systems*. Cambridge University Press, Cambridge, UK.
- [3] A. Bensoussan, M. Çakanyıldırım and S.P. Sethi (2004). *Partially observed inventory systems: The case of zero balance walk*. Working paper SOM 200548, School of Management, University of Texas at Dallas, TX.
- [4] A. Bensoussan, M. Çakanyıldırım and S.P. Sethi (2005). *Optimality of base stock and  $(s, S)$  policies for inventory problems with information delays*. Working paper SOM 200547, School of Management, University of Texas at Dallas, TX.
- [5] A. Bensoussan, M. Çakanyıldırım and S.P. Sethi (2005). *A Multiperiod Newsvendor Problem with Partially Observed Demand*. Working paper SOM 200550, School of Management, University of Texas at Dallas, TX.
- [6] X. Ding, M.L. Puterman and A. Bisi (2002). *The censored newsvendor and the optimal acquisition of information*, Operations Research, Vol.50, No.3: 517-527.
- [7] R.J. Elliott, L. Aggoun and J.B. Moore (1995). *Hidden Markov Models: Estimation and Control*. Springer-Verlag New York, Inc.

- [8] M.L. Fisher, A. Raman and A.S. McClelland (2000). *Rocket science retailing is almost here: Are you ready?* Harvard Business Review, July-August issue: 115-124.
- [9] F.W. Harris (1913). Operations and Cost. *Factory Management Series*, A.-W. Shaw Company, Chicago, Il, Chapter IV: 48-52.
- [10] H.J. Kushner (1967). *Dynamic equations for nonlinear filtering*. Journal of Differential Equations, Vol.3: 179-190.
- [11] A. Müller and D. Stoyan (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley and Sons, New York.
- [12] A. Raman, N. DeHoratius and Z. Ton (2001). *Execution: The missing link in retail operations*. California Management Review Vol.43, No.3: 136-152.
- [13] M. Zakai (1969). *On the optimal filtering of diffusion processes*. Z. Wahrsch. verw. Gebiete, Vol.11: 230-243.