CHAPTER 6

DEALING WITH UNCERTAINTY IN INVENTORY OPTIMIZATION

6.1 INTRODUCTION

The title of this chapter is a little misleading. After all, the inventory models in Chapter 4 deal with uncertainty in inventory optimization, too. But those models assumed a single kind of uncertainty—i.e., demand uncertainty—and assumed that inventory is the only tool for mitigating the uncertainty. In contrast, this chapter uses a broader definition of uncertainty and of the ways that we can mitigate it using inventory. The models in Sections 6.2–6.4 discuss the interactions among multiple inventory locations, and how these locations can pool together—either literally or virtually—to reduce inventory-related costs. Then, in Sections 6.5–6.8, we discuss models for mitigating a different type of uncertainty, namely, supply uncertainty.

6.2 THE RISK-POOLING EFFECT

6.2.1 Overview

Consider a network consisting of N distribution centers (DCs) or other facilities, each of which faces random demand for a single product. The DCs each hold inventory of

Fundamentals of Supply Chain Theory, First Edition. Lawrence V. Snyder, Zuo-Jun Max Shen.

this product. In fact, they act like N independent newsvendors, each facing $N(\mu, \sigma^2)$ demand per period. If the DCs each wish to meet a type-1 service level of α (that is, they wish to stock out in no more than $100(1-\alpha)\%$ of the periods on average), they must each hold an amount of safety stock equal to $z_\alpha \sigma$. The total safety stock in this system is therefore $N z_\alpha \sigma$.

Now suppose that all N DCs are merged into a single DC. What are the inventory implications of this consolidation? (We're ignoring the possible increase in transportation costs and hassle the consolidation may cause.) The new DC's demand process is equal to the sum of all of the original DC's demands. This process has a mean demand of $N\mu$ and a standard deviation of $\sqrt{N}\sigma$. Therefore, to meet the same service level (α) the new DC needs to hold $\sqrt{N} z_\alpha \sigma$ of safety stock, which is less than the safety stock required when N DCs each hold inventory.

This phenomenon is known as the *risk-pooling effect* (Eppen 1979). The basic idea is that by pooling demand streams, we can reduce the amount of safety stock required to meet a given service level, and hence we can reduce the holding cost.

We next discuss the risk-pooling effect in greater generality. Our analysis is adapted from that of Eppen (1979).

6.2.2 Problem Statement

We'll assume that each DC follows a base-stock inventory policy under periodic review, with S_i the base-stock level for DC i. Excess inventory may be stored from period to period (with a holding cost of h per unit per period), and excess demand is backordered (with a penalty cost of p per unit). We assume $p > h$. Note that h and p are the same at every DC.

The demand per period seen by DC i is represented by the random variable D_i, with $D_i \sim N(\mu_i, \sigma_i^2)$. Let f_i and F_i be the pdf and cdf, respectively, of D_i. Demands may be correlated among DCs. The covariance of D_i and D_j is given by σ_{ij} and the correlation coefficient by ρ_{ij}; then $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$.

For each DC, the sequence of events in each period is the same as in Section 4.4.

6.2.3 Decentralized System

We will refer to the system described above as the *decentralized system* since each DC operates independently of the others. S_i is the base-stock level at DC i; this is a decision variable. The expected cost per period at DC i can be expressed as a function of S_i as follows:

$$g_i(S_i) = h \int_0^{S_i} (S_i - d) f_i(d) dd + p \int_{S_i}^{\infty} (d - S_i) f_i(d) dd.$$

This formula is identical to the formula for the newsvendor cost (4.21) except for the subscripts i. Therefore, from Theorems 4.1 and 4.2, the optimal solution is

$$S_i^* = F_i^{-1}\left(\frac{p}{h+p}\right) = \mu_i + z_\alpha \sigma_i,$$

where $\alpha = p/(p+h)$ and z_α is the αth fractile of the standard normal distribution, and the optimal cost at DC i is

$$g_i(S_i^*) = (p+h)\phi(z_\alpha)\sigma_i.$$

(Recall that $\phi(\cdot)$ is the pdf of the standard normal distribution.) Defining $\eta = (p+h)\phi(z_\alpha)$ for convenience, the optimal total expected cost (at all DCs) in the decentralized system, denoted $E[C_D]$, is

$$E[C_D] = \eta \sum_{i=1}^{N} \sigma_i. \tag{6.1}$$

6.2.4 Centralized System

Now imagine that the DCs are consolidated into a single DC that serves all of the demand. We will refer to this as the *centralized system*. Let D_C be the total demand seen by this super-DC. Its mean and standard deviation are

$$\mu_C = \sum_{i=1}^{N} \mu_i$$
$$\sigma_C = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{ij}}.$$

(Note that by definition, $\sigma_{ii} = \sigma_i^2$.) Similar logic as above shows that the optimal base-stock level for the centralized system is

$$S_C^* = \mu_C + z_\alpha \sigma_C$$

with optimal expected cost

$$E[C_C] = \eta\sigma_C = \eta\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{ij}}. \tag{6.2}$$

6.2.5 Comparison

Now let's compare the centralized and decentralized systems. The next theorem says that the centralized system is no more expensive than the decentralized system. This is the risk-pooling effect.

Theorem 6.1 *For the decentralized, N-DC system and the centralized, single-DC system formed by merging the DCs, $E[C_C] \leq E[C_D]$.*

Proof.

$$\begin{aligned}
E[C_C] &= \eta \sqrt{\sum_{i=1}^{N} \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sigma_i \sigma_j \rho_{ij}} \\
&\le \eta \sqrt{\sum_{i=1}^{N} \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sigma_i \sigma_j} \qquad \text{(since } \rho_{ij} \le 1\text{)} \\
&= \eta \sqrt{\left(\sum_{i=1}^{N} \sigma_i \right)^2} \\
&= E[C_D].
\end{aligned}$$

■

One interpretation of the risk-pooling effect is that pooling inventory allows the firm to take advantage of random fluctuations in demand. If one DC sees unusually high demand in a given time period, it's possible that another DC sees unusually low demand. In the centralized system, the excess inventory at the low-demand DC can be used to make up the shortfall at the high-demand DC. In the decentralized system, there is no opportunity for this supply–demand matching.

A more mathematical explanation is that risk pooling occurs because the centralized system takes advantage of the concave nature of safety stock requirements. The amount of safety stock required is proportional to the standard deviation of demand. The standard deviation of demand at the centralized site is smaller than the sum of the standard deviations of the individual sites in the decentralized system since variances, not standard deviations, are additive.

Somewhat surprisingly, the *variances* of the costs of the centralized and decentralized systems are equal at optimality; that is, $\mathrm{Var}[C_D] = \mathrm{Var}[C_C]$, where $\mathrm{Var}[\cdot]$ denotes variance (Schmitt et al. 2010a).

6.2.6 Magnitude of Risk-Pooling Effect

Let's try to get a handle on the magnitude of the risk-pooling effect. Let

$$v = 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sigma_i \sigma_j \rho_{ij}.$$

Note that

$$E[C_C] = \eta \sqrt{\sum_{i=1}^{N} \sigma_i^2 + v}.$$

Uncorrelated Demands: First assume that the demands are uncorrelated, i.e., $\rho_{ij} = 0$ for all $i,\ j$, so $v = 0$. Then

$$E[C_C] = \eta\sqrt{\sum_{i=1}^{N}\sigma_i^2 + v} = \eta\sqrt{\sum_{i=1}^{N}\sigma_i^2} \le \eta\sqrt{\left(\sum_{i=1}^{N}\sigma_i\right)^2} = E[C_D].$$

The magnitude of the difference between $E[C_C]$ and $E[C_D]$ depends on the magnitude between $\sqrt{\sum\sigma_i^2}$ and $\sum\sigma_i$.

Positively Correlated Demands: Next suppose that demands are positively correlated. In fact, consider the extreme case in which $\rho_{ij} = 1$ for all $i,\ j$. Then

$$\begin{aligned} E[C_C] &= \eta\sqrt{\sum_{i=1}^{N}\sigma_i^2 + v} = \eta\sqrt{\sum_{i=1}^{N}\sigma_i^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sigma_i\sigma_j} \\ &= \eta\sqrt{\left(\sum_{i=1}^{N}\sigma_i\right)^2} = \eta\sum_{i=1}^{N}\sigma_i = E[C_D], \end{aligned}$$

so there is no risk pooling effect at all (in the extreme case of perfect correlation).

Negatively Correlated Demands: Finally, assume that demands are negatively correlated. It's difficult to identify the extreme case since ρ_{ij} can't equal -1 for all $i,\ j$. (Why?) But we can say that $v \ge -\sum_{i=1}^{N}\sigma_i^2$ since

$$\sum_{i=1}^{N}\sigma_i^2 + v = \sigma_C^2 \ge 0.$$

So let's assume as an extreme scenario that $v = -\sum_{i=1}^{N}\sigma_i^2$. Then

$$E[C_C] = \eta\sqrt{\sum_{i=1}^{N}\sigma_i^2 - \sum_{i=1}^{N}\sigma_i^2} = 0.$$

The centralized cost is 0, while the decentralized cost is not.

So the risk-pooling effect is very pronounced when demands are negatively correlated, smaller when demands are uncorrelated, and smaller still, or even non-existent, when demands are positively correlated. Why? Recall the explanation given in Section 6.2.5: The risk-pooling effect occurs because excess inventory at one DC can be used to meet excess demand at another. If demands are negatively correlated, there is a lot of opportunity to do this since demands will be very disparate at different locations. On the other hand, if demands are positively correlated, they tend to be all high or all low at the same time, so there is little opportunity for supply–demand matching.

6.2.7 Final Thoughts

The analysis above only considers holding and stockout costs; it does not consider fixed costs (to build and operate DCs) or transportation costs. Clearly, as DCs are consolidated, the fixed cost will decrease. But the transportation cost will increase, since retailers (or other downstream facilities) will be served from more distant DCs. In many cases, the magnitude of the risk pooling effect may be far outweighed by the increases or decreases in fixed and transportation cost. Any analysis of a potential consolidation of DCs must include all factors, not just risk pooling. The location model with risk pooling (LMRP), discussed in Section 8.2, attempts to incorporate all of these factors when choosing facility locations.

6.3 POSTPONEMENT

6.3.1 Introduction

Many firms have product lines containing closely related products. In many cases, multiple end products are made from a single generic product. For example, the clothing retailer Benetton sells many colors of sweater, each of which comes from the same white sweater that's dyed multiple colors (Heskett and Signorelli 1984). Hewlett-Packard sells the same printer in dozens of countries, with a different power supply module, manual, and labels in each (Lee and Billington 1993). IBM builds individualized computers by building partially finished products called "vanilla boxes" and customizing them to order (Swaminathan and Tayur 1998).

A key question in the design of the manufacturing process for each of these products is: When should the end products be differentiated? For example, consider a manufacturer of mobile phones that sells phones in many countries. The company programs each phone with a given language at the factory—the phone is "localized" when it is manufactured. The number of phones to be programmed in each language is determined based on a forecast of the demand in each country. The phones are then shipped to regional distribution centers, approximately one on each continent. The regional DCs store the phones until they are required by retailers, at which point they are shipped to individual countries. If the demand forecasts were wrong, and demand for phones in, say, Thailand was higher than expected while demand in Holland was lower than expected, the company would have to correct this discrepancy by re-programming some of the Dutch phones into Thai phones, then shipping them from the Europe DC to the Asia DC—a costly and time-consuming proposition.

Now suppose that *generic* phones are shipped to the regional DCs, and languages are programmed at the DCs once the phones are requested by retailers. Since the phones are localized on demand, there is much less risk of having too many phones of one language and too few of another. In addition, the firm holds inventory of generic phones, not localized phones, which means that fewer phones need to be held in safety stock due to the risk pooling effect, as we will see below.

This strategy is called *postponement* or *delayed differentiation*. The idea is to delay, as much as possible, the point in the manufacturing process at which end products are differentiated from one another. Of course, designing a postponement strategy may be extremely complicated, since it may require the redesign of the product and the manufacturing and distribution processes. In the mobile phone example, the regional DCs would have to be outfitted with language-programming equipment.

To take the Benetton example to an extreme, postponement might mean that sweaters are dyed in the retail stores once they are demanded by a customer. You would request, say, a red sweater, and it would be dyed for you on demand; stores would never be out of stock of the sweater you wanted. This seems silly, since the costs of implementing such a system would probably far outweigh the benefits. But some products are actually sold this way. For example, paint is mixed to order from generic white paint at your hardware store, giving you access to an enormous range of colors that would be prohibitively expensive to keep in stock. (See Lee (1996) for a discussion of the benefits and challenges of postponement, as well as for two models similar to the model presented below.)

In this section, we will present an analytical model to study the risk-pooling benefit of postponement. This model does not consider the quantitative benefits due to better matching of supply and demand, the improvements in customer satisfaction, or the costs of re-engineering the product and manufacturing process.

6.3.2 Optimization Model

Suppose there are N end products that are made from the same generic product. We will denote the demand for end product i in a given period by D_i and assume that it is normally distributed with a mean of μ_i and a standard deviation of σ_i. For simplicity, we will also assume that demands of different end products are independent of one another, though this assumption is not necessary for the analysis. The manufacturing process takes T periods from the time manufacturing begins to the time the product is sold. (It may be unreasonable to expect this time to be fixed and deterministic, especially in the case of products like Benetton sweaters that sit in inventory for a random amount of time, but we will assume this anyway.) During the first t periods of the manufacturing process, the process is the same for all products; after time t, the manufacturing process is different for each end product. In other words, the product is generic until time t, after which it becomes differentiated. (See Figure 6.1.) In the Benetton example, $t = 0$ might correspond to using dyed wool to produce sweaters (the products are differentiated before manufacturing even begins); $t = T$ corresponds to dyeing the sweaters at the time of sale; and $0 < t < T$ corresponds to an intermediate strategy—for example, dyeing the sweaters after production but before shipping to stores.

Inventory is held of both the generic product (after time t) and each finished product (after time T). Let h_0 be the holding cost per item per period for the generic product and h_i the holding cost for end product i, $i = 1, \ldots, N$. We will assume that $h_0 < h_i$ for all i since value is added as processing continues. We will assume that $t > T/2$ (mainly for mathematical convenience).

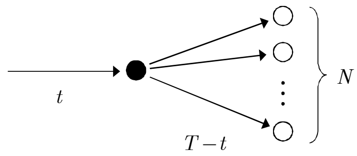

Figure 6.1 Manufacturing process with postponement.

Assuming a desired type 1 service level of α, the required amount of safety stock of the generic product is

$$z_\alpha \sqrt{t} \sqrt{\sum_{i=1}^{N} \sigma_i^2}$$

from (4.44), since the lead time for the generic product is t and the total standard deviation of demand per period is $\sqrt{\sum_i \sigma_i^2}$. The required safety stock of end product i is given by

$$z_\alpha \sqrt{T-t}\sigma_i.$$

Therefore, the total expected holding cost for all products is

$$C(t) = z_\alpha \left[h_0 \sqrt{t} \sqrt{\sum_{i=1}^{N} \sigma_i^2} + \sqrt{T-t} \sum_{i=1}^{N} h_i \sigma_i \right].$$

As t increases, the cost decreases since

$$\begin{aligned} \frac{dC(t)}{dt} &= z_\alpha \left(\tfrac{1}{2} h_0 t^{-\frac{1}{2}} \sqrt{\textstyle\sum \sigma_i^2} - \tfrac{1}{2}(T-t)^{-\frac{1}{2}} \textstyle\sum h_i \sigma_i \right) \\ &< \tfrac{1}{2} z_\alpha h_0 \left(t^{-\frac{1}{2}} \sqrt{\textstyle\sum \sigma_i^2} - (T-t)^{-\frac{1}{2}} \textstyle\sum \sigma_i \right) \\ &< \tfrac{1}{2} z_\alpha h_0 \left(\left[t^{-\frac{1}{2}} - (T-t)^{-\frac{1}{2}} \right] \textstyle\sum \sigma_i \right) \\ &< 0 \end{aligned}$$

since

$$t > T/2 \implies t > T-t \implies \sqrt{t} > \sqrt{T-t} \implies t^{-\frac{1}{2}} - (T-t)^{-\frac{1}{2}} < 0.$$

Therefore, postponement results in decreased costs.

6.3.3 Relationship to Risk Pooling

The cost savings from postponement is due to the risk pooling effect: Generic products represent pooled inventory, while end products represent decentralized inventory.

This relationship can be made explicit by setting $t = 0$ and $t = T$. When $t = 0$, the total safety stock required is

$$z_\alpha \sqrt{T} \sum_{i=1}^{N} \sigma_i,$$

which is proportional to the safety stock required in the decentralized system in our discussion of risk pooling. Similarly, when $t = T$, the total safety stock required is

$$z_\alpha \sqrt{T} \sqrt{\sum_{i=1}^{N} \sigma_i^2},$$

which is proportional to the safety stock in the centralized system.

6.4 TRANSSHIPMENTS

6.4.1 Introduction

When multiple retailers stock the same product, it is sometimes advantageous for one retailer to ship items to another if the former has a surplus and the latter has a shortage. Such "lateral" transfers are called *transshipments*. Transshipments are a mechanism for improving service levels since they allow demands to be satisfied in the current period when they might otherwise be lost or backordered until the following period. In that regard, the benefit from transshipments is very similar to that from risk pooling, since transshipments use one retailer's surplus to reduce another retailer's shortfall. In this case, however, there is no physical pooling of inventory, though the strategy is sometimes referred to as "information pooling." Of course, transshipments come at a cost: Transshipments are often more expensive than replenishments from the DC because they are smaller and therefore lack the economies of scale from larger shipments.

In this section we will discuss a model for setting base-stock levels in a system with two retailers that may transship to one another. This model is adapted from Tagaras (1989). For models with more than two retailers, see Tagaras (1999) or Herer et al. (2006).

This model will assume that transshipments occur after the demand has been realized but before it must be satisfied. Therefore, these transshipments are *reactive* since they are made in reaction to realized demands. In contrast, one might consider *proactive* transshipments that are made in anticipation of demand shortages. Proactive transshipments are of interest when demands must be met instantaneously, since there is no opportunity for transshipping between demand realization and satisfaction. On the other hand, proactive transshipments are more complex to model, so we will focus only on reactive transshipments. We will develop an analytical expression for the expected cost function, but the expected cost can only be minimized using numerical methods (rather than using differentiation). We will also discuss the improvement in service levels due to transshipments.

models that consider uncertainty in supply; in other words, what happens when a firm's suppliers, or the firm's own facilities, are unreliable.

Supply uncertainty may take a number of forms. These include:

- *Disruptions.* A disruption interrupts the supply of goods at some stage in the supply chain. Disruptions tend to be binary events—either there's a disruption or there isn't. During a disruption, there's generally no supply available. Disruptions may be due to bad weather, natural disasters, strikes, suppliers going out of business, etc.

- *Yield Uncertainty.* Sometimes the quantity that a supplier can provide falls short of the amount ordered; the amount actually supplied is random. This is called yield uncertainty. It can be the result of product defects, or of batch processes in which only a certain percentage of a given batch (the yield) is usable.

- *Lead Time Uncertainty.* Uncertainty in the supply lead time can result from stockouts at the supplier, manufacturing or transit delays, and so on. In this case, the lead time L that figures into many of the models in this book must be treated as a random variable rather than a constant.

In this section we will discuss the first two types of supply uncertainty. We will discuss models for setting inventory levels in the presence of disruptions in Section 6.6 and in the presence of yield uncertainty in Section 6.7. In both sections, we will cover models that are analogous to the classical EOQ and infinite-horizon newsvendor models (the models from Sections 3.2 and 4.4.4). Next we discuss the risk-diversification effect, a supply-uncertainty version of the risk-pooling effect.

In most of the models in this section, we will assume that demand is deterministic. We do this for tractability, but also, more importantly, to highlight the effect of supply uncertainty, in the absence of demand uncertainty.

In some ways, there is no conceptual difference between supply uncertainty and demand uncertainty. After all, having too little supply is the same as having too much demand. A firm might use similar strategies for dealing with the two types of uncertainty, as well—for example, holding safety stock, utilizing multiple suppliers, or improving its forecasts of the uncertain events. But, as we will see, the ways in which we model these two types of uncertainty, and the insights we get from these models, can be quite different. (For more on this issue, see Snyder and Shen (2006).)

For reviews of the literature on disruptions, see Snyder et al. (2010) and Vakharia and Yenipazarli (2008), and for yield uncertainty, see Yano and Lee (1995) and Grosfeld Nir and Gerchak (2004). For an overview of models with lead-time uncertainty, see Zipkin (2000).

6.6 INVENTORY MODELS WITH DISRUPTIONS

Disruptions are usually modeled using a two-state Markov process in which one state represents the supplier operating normally and the other represents a disruption.

These states may be known as up/down, wet/dry, on/off, normal/disrupted, and so on. (We'll use the terms up/down.) Not surprisingly, continuous-review models (such as the one in Section 6.6.1) use continuous-time Markov chains (CTMCs) while periodic-review models (Section 6.6.2) use discrete-time Markov chains (DTMCs). The time between disruptions, and the length of disruptions, are therefore exponentially or geometrically distributed (in the case of CTMCs and DTMCs, respectively). The models presented here assume the inventory manager knows the state of the supplier at all times.

Some papers also consider more general disruption processes than the ones we consider here—for example, non-stationary disruption probabilities (Snyder and Tomlin 2006) or partial disruptions (Güllü et al. 1999). These disruption processes can also usually be modeled using Markov processes.

6.6.1 The EOQ Model with Disruptions

6.6.1.1 Problem Statement Consider the classical EOQ model with fixed order cost K and holding cost h per unit per year. The demand rate is d units per year (a change from our notation in Section 3.2). Suppose that the supplier is not perfectly reliable—that it functions normally for a certain amount of time (an up interval) and then shuts down for a certain amount of time (a down interval). The transitions between these intervals are governed by a continuous-time Markov chain (CTMC). During down intervals, no orders can be placed, and if the retailer runs out of inventory during a down interval, all demands observed until the beginning of the next up interval are lost, with a stockout cost of p per lost sale. Both types of intervals last for a random amount of time. Every order placed by the retailer is for the same fixed quantity Q. Our goal is to *choose Q to minimize the expected annual cost*.

This problem, which is known as the *EOQ with disruptions* (EOQD), was first introduced by Parlar and Berkin (1991), but their analysis contained two errors that rendered their model incorrect. A correct model was presented by Berk and Arreola-Risa (1994), whose treatment we follow here.

Let X and Y be the duration of a given up and down interval, respectively. X and Y are exponentially distributed random variables, X with rate λ and Y with rate μ. (Recall that if $X \sim \exp(\lambda)$, then $f(x) = \lambda e^{-\lambda x}$, $F(x) = 1 - e^{-\lambda x}$, and $E[X] = 1/\lambda$.) The parameters λ and μ are called the *disruption rate* and *recovery rate*, respectively. These are the transition rates for the CTMC.

The EOQ inventory curve now looks something like Figure 6.3. Note that the inventory position never becomes negative because excess demands are lost, not backordered. The time between successful orders is called a *cycle*. The length of a cycle, T, is a random variable. If the supplier is in an up interval when the inventory level reaches 0, then $T = Q/d$, otherwise, $T > Q/d$.

Note: In the EOQ, we ignored the per-unit ordering cost c because the annual per-unit cost is independent of Q (since d units are ordered every year, regardless of Q). It is not strictly correct to ignore c in the EOQD because, in the face of lost sales, the number of units ordered each year may not equal d, and in fact it depends on Q. Nevertheless, we will ignore c for tractability reasons.

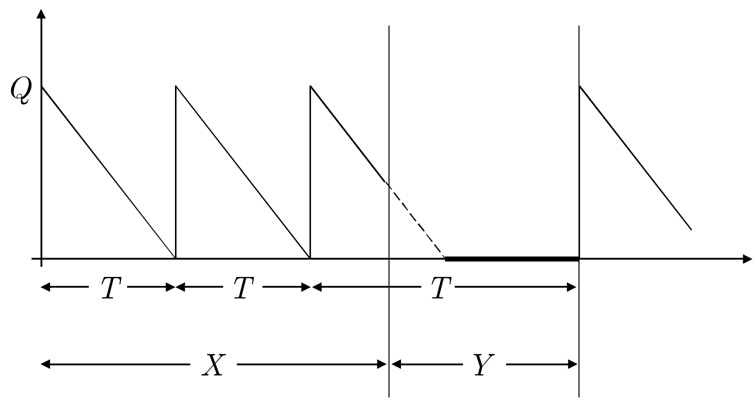

Figure 6.3 EOQ inventory curve with disruptions.

6.6.1.2 Expected Cost Let ψ be the probability that the supplier is in a down interval when the inventory level hits 0. One can show that

$$\psi = \frac{\lambda}{\lambda + \mu}\left(1 - e^{-\frac{(\lambda+\mu)Q}{d}}\right). \tag{6.17}$$

Let $f(t)$ be the pdf of T, the time between successful orders. Then

$$f(t) = \begin{cases} 0, & \text{if } t < Q/d \\ 1-\psi, & \text{if } t = Q/d \\ \psi\mu e^{-\mu(t-Q/d)}, & \text{if } t > Q/d. \end{cases}$$

Note that $f(t)$ has an atom at Q/d and is continuous afterwards.

Each cycle lasts at least Q/d years. After that, with probability $1-\psi$, it lasts an additional 0 years, and with probability ψ, it lasts, on average, an additional $1/\mu$ years (because of the memoryless property of the exponential distribution). Therefore, the expected length of a cycle is given by

$$E[T] = \frac{Q}{d} + \frac{\psi}{\mu}. \tag{6.18}$$

We're interested in finding an expression for the expected cost per year. It's difficult to write an expression for this cost directly. On the other hand, we can calculate the expected cost of one cycle, as well as the expected length of a cycle. Moreover, the system state is always the same at the beginning of each cycle—we have Q units on hand, and the supplier is in an up interval. In situations like these, the well-known *Renewal Reward Theorem* is helpful. (See, e.g., Ross (1995).) In particular, the Renewal Reward Theorem tells us that the expected cost per year, $g(Q)$, is given by

$$g(Q) = \frac{E[\text{cost per cycle}]}{E[\text{cycle length}]}. \tag{6.19}$$

The denominator is given by (6.18); it remains to find an expression for the numerator.

In each cycle we place exactly one order, incurring a fixed cost of K. The inventory in a given cycle is positive for exactly Q/d years (regardless of whether there's a disruption), so the holding cost is based on the area of one triangle in Figure 6.3, namely $Q^2/2d$. Finally, we incur a penalty cost if the supplier is in a down interval when the inventory level hits 0. This happens with probability ψ, and if it does happen, the expected remaining duration of the down interval is $1/\mu$. Therefore, the expected stockout cost per cycle is $pd\psi/\mu$. Then the total expected cost per cycle is

$$K + \frac{hQ^2}{2d} + \frac{pd\psi}{\mu}. \tag{6.20}$$

We can use (6.18)–(6.20) to derive the expected cost per year; the result is stated in the next proposition.

Proposition 6.1 *In the EOQD, the expected cost per year is given by*

$$g(Q) = \frac{K + hQ^2/2d + pd\psi/\mu}{Q/d + \psi/\mu}. \tag{6.21}$$

6.6.1.3 Solution Method Remember that ψ is a function of Q, and in fact it's a pretty messy function of Q. Therefore, (6.21) can't be solved in closed form—that is, we can't take a derivative, set it equal to 0, and solve for Q. Instead, it must be solved numerically using line search techniques such as bisection search. These techniques rely on $g(Q)$ having certain nice properties like convexity. Unfortunately, it is not known whether $g(Q)$ is convex with respect to Q, but it is known that $g(Q)$ is *quasiconvex* in Q. A quasiconvex function has only one local minimum, which is a sufficient condition for most line search techniques to work.

There's nothing wrong with solving the EOQD numerically, insofar as the algorithm for doing so is quite efficient. On the other hand, it's desirable to have a closed-form solution for it for two main reasons. One is that we may want to embed the EOQD into some larger model rather than implementing it as-is. (See, e.g., Qi et al. (2010).) Doing so may require a closed-form expression for the optimal solution or the optimal cost. The other reason is that we can often get insights from closed-form solutions that we can't get from solutions we have to obtain numerically.

Although we can't get an exact solution for the EOQD in closed form, we can get an approximate one. In particular, Snyder (2009) approximates ψ by ignoring the exponential term:

$$\hat{\psi} = \frac{\lambda}{\lambda + \mu}. \tag{6.22}$$

$\hat{\psi}$ is the probability that the supplier is in a down interval at an arbitrary point in time. But ψ refers to a specific point in time, i.e., the point when the inventory level hits 0, and the term $(1 - e^{-(\lambda+\mu)Q/d})$ in the definition of ψ accounts for the knowledge that, when this happens, we were in an up interval as recently as Q/d years ago.

By replacing ψ with $\hat{\psi}$, then, we are essentially assuming that the system approaches steady state quickly enough that when the inventory level hits 0, we can ignore this bit of knowledge, i.e., ignore the transient nature of the system at this moment. The approximation is most effective, then, when cycles tend to be long; e.g., when Q/d is large. If Q/d is large, then $(\lambda+\mu)Q/d$ is large, $e^{-(\lambda+\mu)Q/d}$ is small, and $\hat{\psi} \approx \psi$. The approximation tends to be quite tight for reasonable values of the parameters.

The advantage of using $\hat{\psi}$ in place of ψ is that the resulting expected cost function no longer has any exponential terms, and we can set its derivative to 0 and solve for Q in closed form. This allows us to perform some of the same analysis on the EOQD that we do on the EOQ—for example, we can perform sensitivity analysis, develop worst-case bounds for power-of-two policies, and so on. It also allows an examination of the cost of using the classical EOQ solution when disruptions are possible; as it happens, the cost of this error can be quite large.

6.6.2 The Newsvendor Problem with Disruptions

In this section we consider the infinite-horizon newsvendor problem of Section 4.4.4, except that in place of demand uncertainty we have supply uncertainty, in the form of disruptions. We know from Section 4.4.4 that in the case of demand uncertainty, a base-stock policy is optimal, with the optimal base-stock level given by

$$S^* = \mu + \sigma\Phi^{-1}\left(\frac{p}{p+h}\right) \tag{6.23}$$

(if demand is normally distributed and $\gamma = 1$). We will see that the optimal solution for the problem with supply uncertainty has a remarkably similar form.

The model we discuss below can be viewed as a special case of models introduced by Güllü et al. (1997) and by Tomlin (2006). Elements of our analysis are adapted from Tomlin (2006) and from the unabridged version of that paper (Tomlin 2005). Some of the analysis can also be found in Schmitt et al. (2010b).

6.6.2.1 Problem Statement As in Section 6.6.1 on the EOQD, we assume that demand is deterministic; it's equal to d units per period. (d need not be an integer.) On-hand inventory and backorders incur costs of h and p per unit per period, respectively. There is no lead time. The sequence of events is identical to that described in Section 4.4, except that in step 2, no order is placed if the supplier is disrupted.

The probability that the supplier is disrupted in the next period depends on its state in the current period. In other words, the disruption process follows a two-state discrete-time Markov chain (DTMC). Let

$$\begin{aligned} \alpha &= P(\text{down next period}|\text{up this period}) \\ 1-\beta &= P(\text{down next period}|\text{down this period}). \end{aligned}$$

We refer to α as the *disruption probability* and β as the *recovery probability*. These are the transition probabilities for the DTMC. The up and down periods both constitute geometric processes, and these processes are the discrete-time analogues to the continuous-time up/down processes in Section 6.6.1.

Given the transition probabilities α and β, we can solve the Chapman-Kolmogorov equations to derive the steady-state probabilities of being in an up or down state as:

$$\pi_u = \frac{\beta}{\alpha + \beta} \tag{6.24}$$

$$\pi_d = \frac{\alpha}{\alpha + \beta} \tag{6.25}$$

It turns out to be convenient to work with a more granular Markov chain that indicates not only whether the supplier is in an up or down period, but also how long the current down interval has lasted. In particular, state n in this Markov chain represents being in a down interval that has lasted for n consecutive periods. If $n = 0$, we are in an up period.

Let π_n be the steady-state probability that the supplier is in a disruption that has lasted n periods. Furthermore, define

$$F(n) = \sum_{i=0}^{n} \pi_n. \tag{6.26}$$

$F(n)$ is the cdf of this process and represents the steady-state probability that the supplier is in a disruption that has lasted n periods or fewer (including the probability that it is not disrupted at all). These probabilities are given explicitly in the following lemma, but often, we will ignore the explicit form of the probabilities and just use π_n and $F(n)$ directly.

Lemma 6.1 *If the disruption probability is α and the recovery probability is β, then*

$$\begin{aligned}
\pi_0 &= \frac{\beta}{\alpha + \beta} \\
\pi_n &= \frac{\alpha\beta}{\alpha + \beta}(1-\beta)^{n-1}, \quad n \geq 1 \\
F(n) &= 1 - \frac{\alpha}{\alpha + \beta}(1-\beta)^{n}, \quad n \geq 0.
\end{aligned}$$

Proof. Omitted; see Problem 6.14. ∎

6.6.2.2 Form of the Optimal Policy Our objective is to make inventory decisions to minimize the expected holding and stockout cost per period. What type of inventory policy should we use? It turns out that a base-stock policy is optimal for this problem:

Theorem 6.5 *A base-stock policy is optimal in each period of the infinite-horizon newsvendor problem with deterministic demand and stochastic supply disruptions.*

We omit the proof of Theorem 6.5; it follows from a much more general theorem proved by Song and Zipkin (1996). Note that a base-stock policy works somewhat differently in this problem than in previous problems, since we might not be able to order up to the base-stock level in every period—in particular, we can't order *anything* during down periods. So a base-stock policy means that we order up to the base-stock level during up periods and order nothing during down periods. The extra inventory during up periods is meant to protect us against down periods.

6.6.2.3 Expected Cost Suppose the supplier is in state $n = 0$; that is, an up period. If we order up to a base-stock level of S at the beginning of the period, we incur a cost at the end of the period of

$$h(S-d)^+ + p(d-S)^+. \tag{6.27}$$

In state $n = 1$, we incur a cost of

$$h(S-2d)^+ + p(2d-S)^+, \tag{6.28}$$

and in general, we incur a cost of

$$h\left[S-(n+1)d\right]^+ + p\left[(n+1)d-S\right]^+ \tag{6.29}$$

in state n, for $n = 0, 1, \ldots$.

Therefore, the expected holding and stockout costs per period can be expressed as a function of S as follows:

$$g(S) = \sum_{n=0}^{\infty} \pi_n \left[h\left[S-(n+1)d\right]^+ + p\left[(n+1)d-S\right]^+ \right]. \tag{6.30}$$

In addition, we can say the following:

Lemma 6.2 *The optimal base-stock level S^* is an integer multiple of d.*

Proof (sketch). The proof follows from the fact that g is a piecewise-linear function of S, with breakpoints at multiples of d. ■

Normally, we would find the optimal S by taking a derivative of $g(S)$, but since S is discrete (by Lemma 6.2), we need to use a *finite difference* instead. A finite difference is very similar to a derivative except that, instead of measuring the change in the function as the variable changes infinitesimally, it measures the change as the variable changes by one unit. In particular, S^* is the smallest S such that $\Delta g(S) \geq 0$, where

$$\Delta g(S) = g(S+d) - g(S). \tag{6.31}$$

(Normally, we would define $\Delta g(S)$ as $g(S+1) - g(S)$, but since S can only take on values that are multiples of d, it's sufficient to define $\Delta g(S)$ as in (6.31).)

$$\Delta g(S) = g(S+d) - g(S)$$

$$= \sum_{n=0}^{\infty} \pi_n \Big[h\,[S-nd]^+ + p\,[nd-S]^+ \\ -h\,[S-(n+1)d]^+ - p\,[(n+1)d-S]^+ \Big]$$

Now,

$$[S-nd]^+ - [S-(n+1)d]^+ = \begin{cases} d, & \text{if } n < \frac{S}{d} \\ 0, & \text{otherwise} \end{cases}$$

and

$$[nd-S]^+ - [(n+1)d-S]^+ = \begin{cases} -d, & \text{if } n \geq \frac{S}{d} \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$\begin{aligned} \Delta g(S) &= d \left[h \sum_{n=0}^{\frac{S}{d}-1} \pi_n - p \sum_{n=\frac{S}{d}}^{\infty} \pi_n \right] \\ &= d \left[hF\left(\frac{S}{d}-1\right) - p\left(1 - F\left(\frac{S}{d}-1\right)\right)\right] \\ &= d \left[(h+p)F\left(\frac{S}{d}-1\right) - p \right], \end{aligned}$$

where F is as defined in (6.26). Then S^* is the smallest multiple of d such that

$$(h+p)F\left(\frac{S}{d}-1\right) - p \geq 0 \tag{6.32}$$

$$\iff S \geq d + dF^{-1}\left(\frac{p}{p+h}\right). \tag{6.33}$$

The notation in (6.33) is a little sloppy since $F^{-1}(\gamma)$ only exists if γ happens to be one of the discrete values that $F(n)$ can take. If $p/(p+h)$ is not one of these values, then (6.32) implies it is always optimal to "round up." Interpreted this way, $F^{-1}(\gamma)$ is an integer for all γ, the right-hand side of (6.33) is automatically a multiple of d, and we can drop the "smallest multiple of d" language and replace the inequality in (6.33) with an equality.

We have now proved the following:

Theorem 6.6 *In the infinite-horizon newsvendor problem with deterministic demand and stochastic supply disruptions, the optimal base-stock level is given by*

$$S^* = d + dF^{-1}\left(\frac{p}{p+h}\right), \tag{6.34}$$

where F is as defined in (6.26).

Notice that the optimal base-stock level under supply uncertainty has a very similar structure to that under demand uncertainty, as given in (6.23). First, it uses the familiar newsvendor critical fractile $p/(p + h)$, but here the inverse cdf F refers not to the demand distribution but to the supply distribution.

Second, the right-hand side of (6.34) has a natural cycle stock–safety stock interpretation, just like in the demand uncertainty case. Here, d is the cycle stock—the inventory to meet this period's demand—and $dF^{-1}(\gamma)$, where $\gamma = p/(p + h)$, is the safety stock—the inventory to protect against uncertainty (in this case, supply uncertainty).[1]

Just like in the demand-uncertainty case, the optimal solution specifies what fractile of the distribution we should protect against. Here, we should have enough inventory to protect against any disruption whose length is no more than $F^{-1}(\gamma)$ periods. The probability of a given period being in a disruption that has lasted longer than this is $1 - \gamma$, so, as in the demand-uncertainty case, the type-1 service level is given by γ. As usual, the base-stock level increases with p and decreases with h.

6.7 INVENTORY MODELS WITH YIELD UNCERTAINTY

In some cases, the number of items received from the supplier may not equal the number ordered. This may happen because of stockouts or machine failures at the supplier, or because the production process is subject to defects. The quantity actually received is called the *yield*. If the yield is deterministic—e.g., we always receive 80% of our order size—then the problem is easy: we just multiply our order size by $1/0.8 = 1.25$. More commonly, however, there is a significant amount of uncertainty in the yield. The optimal solution under *yield uncertainty* generally involves increasing the order quantity, as under imperfect but deterministic yield, but it should account for the variability in yield, not just the mean—just as in the case of demand uncertainty.

In the sources of yield uncertainty mentioned above, we'd expect that the actual yield should always be less than or equal to the order quantity—we shouldn't receive more than we order. But yield uncertainty can also occur in batch production processes—e.g., for chemicals or pharmaceuticals. In this case, it's not a matter of items being "defective," but rather of not knowing in advance precisely how much usable product will result from the process. In this case, the amount received may be more than the amount expected, and we can't necessarily place an upper bound on the yield.

In this section, we consider how to set inventory levels under yield uncertainty. As in Section 6.6, we consider both a continuous-review setting, based on the EOQ model, and a periodic-review setting, based on the newsvendor problem. As before, we will assume that demand is deterministic.

There are many ways to model yield uncertainty. We will consider two that are intuitive and tractable.

The first is an *additive yield* model in which we assume that if an order of size Q is placed, then the yield (the amount received) equals $Q + Y$. Y is a continuous

[1] In earlier chapters we used $\alpha = p/(p + h)$; here we use γ since α has a new meaning in this section.

Theorem 6.7 *For the decentralized N-DC system with supply disruptions and deterministic demand, and the centralized, single-DC system formed by merging the DCs:*

1. $S^*_C = NS^*_D = NS^*_C$
2. $E[C_C] = E[C_D] = NE[C]$
3. $\mathrm{Var}[C_C] = N\mathrm{Var}[C_D] = N^2\mathrm{Var}[C]$

6.8.5 Supply Disruptions and Stochastic Demand

Suppose now that demand is uncertain, as in Section 6.2. Disruptions are also still present, as in the preceding analysis.

Under demand uncertainty, the risk-pooling effect says that centralization is preferable, while under supply uncertainty, the risk-diversification effect says that decentralization is preferable. So, if both types of uncertainty are present, which strategy is better? We cannot answer this question analytically since the expected cost function cannot be optimized in closed form for either system. Instead, we evaluate the question numerically.

Most decision makers are risk averse—they are willing to sacrifice a certain amount of expected cost in order to reduce the variance of the cost. One way of modeling risk aversion is using a *mean–variance objective*, popularized by Markowitz in the 1950s:

$$(1-\kappa)E[C] + \kappa\mathrm{Var}[C], \tag{6.48}$$

where $\kappa \in [0, 1]$ is a constant. If κ is small, then the decision maker is fairly risk neutral; the larger κ is, the more risk-averse the decision maker is. Typically κ is less than, say, 0.05.

One can write out $E[C]$ and $\mathrm{Var}[C]$ for the systems with disruptions and demand uncertainty, but we omit the formulas here. Schmitt et al. (2010a) perform a computational study to determine which system is preferable to the risk-averse decision maker. They numerically optimize (6.48) for both the centralized and decentralized systems and determine which system gives the smaller optimal objective value.

They find that the decentralized system is almost always optimal, i.e., that the risk-diversification effect almost always trumps the risk-pooling effect. For example, under a given set of problem parameters, the decentralized system is optimal whenever $\kappa \geq 0.0008$ and $p/(p+h) \geq 0.5$—in other words, whenever the decision maker is even slightly risk averse and the required service level is at least 50%.

PROBLEMS

6.1 (Risk-Pooling Example) Three distribution centers (DCs) each face normally distributed demands, with $D_1 \sim N(22, 8^2)$, $D_2 \sim N(19, 4^2)$, and $D_3 \sim N(17, 3^2)$. All three DCs have a holding cost of $h = 1$ and $p = 15$, and all three follow a periodic-review base-stock policy using their optimal base-stock levels.

a) Calculate the expected cost of the decentralized system.

b) Suppose demands are uncorrelated among the three DCs: $\rho_{12} = \rho_{13} = \rho_{23} = 0$. Calculate the expected cost of the centralized system.

c) Suppose $\rho_{12} = \rho_{13} = \rho_{23} = 0.75$. Calculate the expected cost of the centralized system.

d) Suppose $\rho_{12} = 0.75$, $\rho_{13} = \rho_{23} = -0.75$. Calculate the expected cost of the centralized system.

6.2 (No Soup for You) A certain New York City soup vendor sells 15 varieties of soup. The number of customers who come to the soup store on a given day has a Poisson distribution with a mean of 250. A given customer has an equal probability of choosing each of the 15 varieties of soup, and if his or her chosen variety of soup is out of stock (no pun intended), he or she will leave without buying any soup.

You may assume (although it is not necessarily a good assumption) that the demands for different varieties of soup are independent; that is, if the demand for variety i is high on one day, that doesn't indicate anything about the demand for variety j.

Every type of soup sells for $5 per bowl, and the ingredients for each bowl of soup cost the soup vendor $1. Any soups (or ingredients) that are unsold at the end of the day must be thrown away.

a) How many ingredients of each variety of soup should the soup vendor buy? What is the restaurant's total expected underage and overage cost for the day?

b) What is the probability that the vendor stocks out of a given variety of soup?

c) Now suppose that the soup vendor wishes to streamline his offerings by reducing the selection to 8 varieties of soup. Assume that the total demand distribution does not change, but now the total demand is divided among 8 soup varieties instead of 15. As before, assume that a customer finding his or her choice of soup unavailable will leave without purchasing anything. Now how many ingredients of each variety of soup should the vendor buy? What is the restaurant's total expected underage and overage cost for the day?

d) In a short paragraph, explain how this problem relates to risk pooling.

Note: You may use the normal approximation to the Poisson distribution, but make sure to specify the parameters you are using.

6.3 (Mile-High Trash) On a certain airline, the flight attendants collect trash during flights and deposit it all into a single receptacle. Airline management is thinking about instituting an on-board recycling program in which waste would be divided by the flight attendants and placed into three separate receptacles: one for paper, one for cans and bottles, and one for other trash.

The volume of each of the three types of waste on a given flight is normally distributed. The airline would maintain a sufficient amount of trash-receptacle space on each flight so that the probability that a given receptacle becomes full under the new system is the same as the probability that the single receptacle becomes full under the old system.

Would the new policy require the same amount of space, more space, or less space for trash storage on each flight? Explain your answer in a short paragraph.

6.4 (Days-of-Supply Policies) Rather than setting safety stock levels using base-stock or (r, Q) policies, some companies set their safety stock by requiring a certain number of "days of supply" to be on hand at any given time. For example, if the daily demand has a mean of 100 units, the company might aim to keep an extra 7 days of supply, or 700 units, in inventory. This policy uses μ instead of σ to set safety stock levels.

Consider the N-DC system described in Section 6.2.1, with independent demands across DCs ($\rho_{ij} = 0$ for $i \neq j$). You may assume that all DCs are identical: $\mu_i = \mu$ and $\sigma_i = \sigma$ for all i. Assume that μ and σ refer to *weekly* demands, and that orders are placed by the DCs once per week. Finally, assume that each DC follows a days-of-supply policy with k days of supply required to be on hand as safety stock; each DC's order-up-to level is then

$$S = \mu + \frac{k}{7}\mu.$$

- **a)** Prove that the centralized and decentralized systems have the *same amount* of total inventory.
- **b)** Derive expressions for $E[C_D]$ and $E[C_C]$, the total expected costs of the decentralized and centralized systems. Your expressions may *not* involve integrals; they *may* involve the standard normal loss function, $\mathscr{L}(\cdot)$.

  Hint: Since the DCs are not following the optimal stocking policy, the cost is analogous to (4.35), not to (4.37).
- **c)** Prove that $E[C_C] < E[C_D]$.
- **d)** Explain in words how to reconcile parts (a) and (c)—how can the centralized cost be smaller even though the two systems have the same amount of inventory?

6.5 (Negative Safety Stock) Consider the N-DC system described in Section 6.2.1, with independent demands across DCs ($\rho_{ij} = 0$). Suppose that the holding cost is greater than the penalty cost: $h > p$.

- **a)** Prove that *negative* safety stock is required at DC i—that the base-stock level is less than the mean demand.
- **b)** Prove that the total inventory (cycle stock and safety stock) required in the decentralized system (each DC operating independently) is *less* than the total inventory required in the centralized system (all DCs pooled into one). (This is the opposite of the result in Section 6.2.)
- **c)** Prove that, despite the result from part (b), the total expected cost of the centralized system is less than that of the decentralized system ($E[C_C] < E[C_D]$).
- **d)** Explain in words how to reconcile parts (b) and (c)—how can it be less expensive to hold more inventory?

6.6 (Rationalizing DVR Models) A certain brand of digital video recorder (DVR) is available in three models, one that holds 40 hours of TV programming, one that holds 80 hours, and one that holds 120 hours. The lifecycle for a given DVR model is short, roughly 1 year. Because of long manufacturing lead times, the company must manufacture all of the units it intends to sell before the DVRs go on the market, and it will not have another opportunity to manufacture more before the end of the products' 1-year life cycles.

Demand for DVRs is highly volatile, and customers are very picky. A customer who wants a given model but finds that it's out of stock will almost never change to a different model—instead, he or she will buy a competitor's product. In this case, the firm incurs both the lost profit and a loss-of-goodwill cost. Moreover, any DVRs that are unsold at the end of the year are taken off the market and destroyed, with no salvage value (or cost).

The three models have the following parameters:

Storage Space	Manufacturing Cost (c_i)	Selling Price (r_i)	Goodwill Cost (g_i)	Mean Annual Demand (μ_i)	SD of Annual Demand (σ_i)
40	80	120	150	40,000	12,000
80	90	150	150	55,000	15,000
120	100	250	150	25,000	8,000

Demands are normally distributed with the parameters specified in the table. Moreover, demands for the 80- and 120-hour models are negatively correlated, with a correlation coefficient of $\rho_{80,120} = -0.4$. (Demands for the 40-hour model are independent of those for the other two models.)

The company is currently designing its three models for next year, and a very smart supply chain manager noticed that although the models sell for different prices, they cost nearly the same amount to manufacture. The manager thus proposed that the firm manufacture only a *single* model, containing 120 hours of storage space. When customers purchase a DVR, they specify how much storage space they'd like it to have (either 40, 80, or 120 hours) and pay the corresponding price, and the unit is activated with that much space. If the customer asks for 40 or 80 hours, the remaining storage space simply goes unused. This change can be made with software rather than hardware and therefore costs very little to make.

a) Let Q_i be the quantity of model i manufactured, $i = 1, \ldots, 3$, if the supply chain manager's proposal is *not* followed. Write the firm's expected profit for model i as a function of Q_i.

b) Find the optimal order quantities Q_i^* and the corresponding total optimal expected profit (for all three models).

c) Let Q be the quantity of the single model manufactured if the manager's proposal *is* followed. Write the firm's total expected profit as a function of Q. Although it is not entirely accurate to do so, you may assume that the expected selling price for the single model is given by a weighted average of the r_i, with weights given by the μ_i.

d) Find the optimal order quantity Q and the corresponding optimal expected profit. Based on this analysis, should the firm follow the manager's suggestion?

e) What other factors should the firm consider before deciding whether to implement the manager's proposal?

6.7 (Proof of Theorem 6.3) Prove Theorem 6.3.

6.8 (Transhipment Simulation) Build a spreadsheet simulation model for the two-retailer transshipment problem from Section 6.4. Your spreadsheet should include columns for the demand at each location; the inventory at each location at the start of the period, before transshipments, and after transshipments; the amount transshipped; and the costs for the period. Assume that demands are Poisson with mean λ_i per period and that

$$
\begin{aligned}
\lambda_1 &= 30 & \lambda_2 &= 20 \\
c_1 &= 1.2 & c_2 &= 1.7 \\
h_1 &= 0.6 & h_2 &= 0.8 \\
p_1 &= 8.0 & p_2 &= 8.0 \\
c_{12} &= 3.0 & c_{21} &= 3.0.
\end{aligned}
$$

Use $S_1 = 33$ and $S_2 = 22$ as the base-stock levels, and assume that both retailers begin the simulation with $S_i - \lambda_i$ units on-hand (that is, at the start of period 1, retailer i needs to order λ_i units to bring its inventory position to S_i).

a) Simulate the system for 500 periods and include the first 10 rows of your spreadsheet in your report.

b) Compute the average ordering, transshipment, holding, and penalty costs per period from your simulation.

c) Compute the expected transshipment quantity from retailer 1 to retailer 2 ($E[Y_{12}]$) and the expected ending inventory at retailer 1 ($E[IL_1^+]$) using (6.8) and (6.9). To compute these quantities, you will need to evaluate some integrals numerically.

d) Compare the results from parts (a) and (c). How closely do the simulated and actual quantities match?

e) By trial and error, try to find the values of S_1 and S_2 that minimize the simulated cost. What are the optimal values, and what is the optimal expected cost?

6.9 (Binary Transshipments) Consider the transshipment model from Section 6.4, except now suppose the demands are binary. That is, the demands can only equal 0 or 1, and they are governed by a Bernoulli distribution: $D_i = 1$ with probability q_i and $D_i = 0$ with probability $1 - q_i$, for $i = 1, 2$. All of the remaining assumptions from Section 6.4.2 hold.

Your goal in this problem will be to formulate the expected cost and evaluate several feasible values for the base-stock levels (S_1, S_2). Assume that S_i must be an integer.

a) Explain why $S_1^* + S_2^* \leq 2$.

b) For each possible solution (S_1, S_2) below, write the expected values of the state variables Q_i, Y_{ij}, IL_i^+, and IL_i^-, and then write the expected cost $g(S_1, S_2)$.

1. $(S_1, S_2) = (0, 0)$
2. $(S_1, S_2) = (1, 1)$
3. $(S_1, S_2) = (1, 0)$
4. $(S_1, S_2) = (2, 0)$

(The cases in which $(S_1, S_2) = (0, 1)$ or $(0, 2)$ are similar to the cases above, so we'll skip them.)

Hint 1: If $S_i = 0$, that does *not* mean that stage i never orders!

Hint 2: To check your cost functions, we'll tell you the following: If $c_i = h_i = p_i = 1$, $c_{ij} = 3$, and $q_i = 0.5$ for all $i = 1, 2$, then $g(0, 0) = 4$, $g(1, 1) = 2$, $g(1, 0) = 2.25$, and $g(2, 0) = 3$. Note, however, that these parameters do not satisfy the assumptions on page 152.

c) Prove that, if $h_i \leq p_i$ and $q_i \geq 0.5$ for $i = 1, 2$, and if the assumptions on page 152 are satisfied, then $g(0, 0) \geq g(1, 1)$.

d) Find an instance for which $(S_1^*, S_2^*) = (1, 1)$. Your instance must satisfy the assumptions on page 152.

e) Find a *symmetric* instance for which $(S_1^*, S_2^*) = (1, 0)$. Your instance must satisfy the assumptions on page 152. A symmetric instance is one for which the parameters for the two retailers are identical ($c_1 = c_2$, $h_1 = h_2$, etc.). (It's a little surprising that a symmetric instance can produce a non-symmetric solution, but it can.)

f) Prove or disprove the following claim: $g(2, 0) \geq g(1, 1)$ for all instances that satisfy the assumptions on page 152.

6.10 (EOQD Approximation) Suppose that, in the EOQD model of Section 6.6.1, we replace ψ (a function of Q) with

$$\hat{\psi} = \frac{\lambda}{\lambda + \mu}$$

(which is independent of Q). Let $\hat{g}$ be the cost function that results from replacing ψ with $\hat{\psi}$ in (6.21). It is known that $\hat{g}$ is convex (you do not need to prove this).

a) Prove that the derivative of $\hat{g}(Q)$ is

$$\hat{g}'(Q) = \frac{\frac{h\mu^2}{2}Q^2 + \psi dh\mu Q - (Kd\mu + d^2 p\psi)\mu}{(Q\mu + \psi d)^2}.$$

b) Prove that $\hat{Q}^*$, the Q that minimizes $\hat{g}$, is given by

$$\hat{Q}^* = \frac{-\psi dh + \sqrt{(\psi dh)^2 + 2hd\mu(K\mu + dp\psi)}}{h\mu}. \tag{6.49}$$

6.11 (Implementing EOQD Approximation) Consider an instance of the EOQD with $K = 35$, $h = 4$, $p = 22$, $d = 30$, $\lambda = 1$, and $\mu = 12$.

a) Find Q^* for this instance using optimization software of your choice. Report the expected cost, $g(Q^*)$.

b) Consider the following heuristic for the EOQD:

1. Set Q equal to the EOQ.
2. Calculate ψ using the current value of Q.
3. Find Q using (6.49) from Problem 6.10, setting $\hat{\psi}$ equal to the current ψ from step 2.
4. If Q has changed more than ϵ since the previous iteration (for fixed $\epsilon > 0$), then go to 2; otherwise, stop.

Using this heuristic and any software package you like, find a near-optimal Q using $\epsilon = 10^{-3}$. Report the Q you found, its cost $g(Q)$, and the percentage difference between $g(Q)$ and $g(Q^*)$ from part (a).

6.12 (Disruption-Prone Bicycle Parts) A bicycle manufacturer buys a particular cable used in its bicycles from a single supplier located in South America. The manufacturer follows a periodic-review base-stock policy, placing an order with the supplier every week. The supplier occasionally experiences disruptions due to hurricanes, labor actions, and other events. These disruptions follow a Markov process with disruption probability $\alpha = 0.1$ and recovery probability $\beta = 0.4$. When not disrupted, the supplier's lead time is negligible. Cables are used by the manufacturer at a constant rate of 6000 per week. Inventory incurs a holding cost of \$0.002 per cable per week. If the manufacturer runs out of cables, it must delay production, resulting in a cost that amounts to \$0.05 per cable per week.

a) On average, how many weeks per year is the supplier disrupted? On average, how long does each disruption last?

b) What is the optimal base-stock level for cables?

6.13 (Optimal Cost for Base-Stock Policy with Disruptions) Prove that, in the base-stock problem with disruptions discussed in Section 6.6.2, the optimal cost is given by

$$g(S^*) = d\left[p \sum_{n=R+1}^{\infty} \pi_n n - h \sum_{n=0}^{R} \pi_n n\right],$$

where $R = F^{-1}(p/(p+h))$ and $F(x)$ is as defined in (6.26). You may assume that h and p are set so that $p/(p+h)$ exactly equals one of the possible values of $F(x)$.

6.14 (Proof of Lemma 6.1) Prove Lemma 6.1.

6.15 (Disruptions = Stochastic Demand?)

a) Develop a *stochastic demand process* that is equivalent to the *stochastic supply process* in the base-stock model with disruptions from Section 6.6.2.

In particular, formulate a demand distribution such that, if the demand is iid stochastic following your distribution but the supply is deterministic, the expected cost is equal to the expected cost given by (6.30), assuming we order up to the same S in every period. Prove that the two expected costs are equal. Make sure you specify both the possible values of the demand and the probability of each value, i.e., the pmf.

b) In part (a) you proved that, under the optimal solution, the expected cost is the same in both models. Is the entire distribution of the random variable representing the cost also the same in both models?

6.16 **(Random Yield for Steel)** Return to Problem 3.1. Suppose that the amount of steel delivered by the supplier differs randomly from the order quantity, and the auto manufacturer must accept whatever quantity the supplier delivers. Let Q be the order quantity.

a) Suppose the delivery quantity is given by $Q + Y$, where $-Y \sim \exp(0.02)$. What is Q^*?

b) Suppose the delivery quantity is given by QZ, where $Z \sim U[0.8, 1.0]$. What is Q^*?

6.17 **(Staffing Truck Drivers)** The U.S. trucking industry suffers from notoriously high employee turnover, with turnover rates often well in excess of 100% (Paz-Frankel 2006). This makes advance planning difficult since it is difficult to predict how many drivers will be available when needed. Suppose a trucking company needs 25 drivers every day. If the company asks S drivers to report to work on a given day, the number of drivers who actually show up is given by $S + Y$, where $Y \sim U[-5, 0]$. Drivers who report to work but are not needed must still be paid their daily wage of \$150. For each driver fewer than 25 that show up, the company will be unable to deliver a load, incurring a cost of \$1200. Find S^*, the optimal number of drivers to ask to report to work.

FUNDAMENTALS OF SUPPLY CHAIN THEORY

Lawrence V. Snyder
Lehigh University

Zuo-Jun Max Shen
University of California, Berkeley

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION