Chapter 14

Inventory Management: Information, Coordination and Rationality[1]

Özalp Özer

Management Science and Engineering
Stanford University
Stanford, CA 94305
oozer@stanford.edu

Abstract

The success of a product in today's global marketplace depends on capabilities of firms in the product's supply chain. Among these capabilities, effective inventory management is a capability necessary to lead in the global marketplace. The chapter provides a discussion of four fundamentals of effective inventory management. First, it requires managers to know how best to use available information. Second, managers need to quantify the value of information. Third, they need to coordinate decentralized inventory operations. Finally, effective inventory management requires decision tools that can be embraced by their users. The chapter's emphasis is on the use of information, and the role of new information technologies in inventory management. Previous research on inventory management played an important role in the advancement and development of new technologies and processes. Today more research is needed because new technologies (such as RFID Radio-Frequency Identification) and new management methods (such as collaborative forecasting and planning) are emerging and evolving faster than ever before. Inventory management and research will continue to play a central role in the success of a product and the firms in its supply chain. The chapter brings together separate but inherently related streams of research in inventory management. By doing so, we highlight potential research opportunities that lie on the boundaries.

---

[1]This manuscript will appear in the Handbook of Production Planning. (Eds) K. Kempf, P. Keskinocak and R. Uzsoy.

1   Introduction

Inventory control problems have attracted researchers for many years[2]. Fundamentally, the problem is one of matching supply and demand by efficiently coordinating the production and the distribution of goods. Recent developments in information technology have equipped managers with the means to obtain better and timely information regarding, for example, demand, lead times, available assets and capacity. Technology has also enabled customers to obtain vast amounts of information about a product, such as its physical attributes and availability. In today's increasingly competitive marketplace, consumers are constantly pressuring suppliers to simultaneously reduce costs and lead times and increase the quality of their products. Good inventory management is no longer a competitive advantage. It is an essential capability to survive in a global market.

An important aspect of good inventory management is effective use of information. Knowing how to use information effectively also enables a manager to decide what data to collect, buy and store, and what information technology to invest in. Note that information has no value, if it is not used effectively. For example, an inventory manager can obtain order progress information through the use of a tracking technology. If this information is not used to improve replenishment decisions, then neither the information nor the technology used to obtain it has any value. In this chapter, we provide some examples of how information is incorporated into classical inventory management problems.

The second important aspect of good inventory management is to quantify the value of information. A manager may need to invest in a technology that collects and stores information relevant for effective inventory management. The cost of obtaining information is often not difficult to analyze. Quantifying the benefits, however, requires thorough analysis and modeling. Consider, for example, the recent tracking technology known as Radio Frequency Identification (RFID). Quantifying the cost of RFID implementation is relatively straightforward. But the benefit of this technology for the management of inventory is not clear. Comparing inventory models with and without the information obtained through RFID enables an inventory manager to quantify the value of RFID. In this chapter, we provide modeling examples through which an inventory manager can quantify the value of information.

The third important aspect of good inventory management is to coordinate decentralized operations. The coordination of information and inventory management have become increasingly more difficult with recent increases in supply chain complexity. Such complexities are the result of dramatic changes in manufacturing and distribution, including globalization and outsourcing. As a result, independent firms manage inventory allocated across different parts of the global supply chains. Each firm in the supply chain individually and myopically sets strategic and operational goals to minimize inventory and production related costs. Firms also maximize profits by using local information such as local cost structures, profit margins and forecasts. As a result, the supply chain is sub-optimized and not synchronized.

We have observed in the past that inability to optimize and synchronize these very complex inventory management issues can lead to catastrophic supply chain failures that make top business news. In 2001, Solectron, a major electronics manufacturer, had $4.7 billion in excess component

---

[2]Throughout the chapter we use the terms inventory/production control, replenishment/production and order/produce interchangeably.

capacity due to inflated forecasts provided by its customers. For exactly the same reason, Cisco, a major telecommunication equipment manufacturer, held \$2.1 billion in excess inventory during the same year. Anticipating such inflation, manufacturers may discount the forecast information. Unfortunately, this caution, e.g., second guessing the forecast, may also lead to huge losses. In 1997, Boeing's suppliers were unable to fulfill Boeing's large orders because they did not believe in Boeing's forecasts. In this chapter we provide examples of research that show such catastrophic outcomes are due to misaligned incentives and lack of coordination. These research works consider the interaction among multiple inventory managers and illustrate how these managers can align incentives through structured agreements and avoid (or mitigate) the adverse effects of lack of coordination.

Finally, good inventory management requires decision tools that can be embraced by their users. The formulations and the methodologies developed in multi-echelon production and distribution systems are often very difficult to explain to non-mathematically oriented students and practitioners. In addition, data fed to these tools are not always accurate. Systems and people are bounded by limited information. In this chapter, we provide a discussion of some efforts to efficiently control multi-period, multi-product supply chains by developing easy-to-describe, near-optimal and robust heuristics that can be implemented on a spreadsheet by solving, for example, newsvendor type problems.

To summarize, the chapter aims to provide a discussion of various topics and concepts from the centralized and decentralized inventory management literature. The emphasis will be on the use of information, and the role of new information technologies in inventory management. We provide examples of some ongoing research work. Our focus is on the modeling aspect rather than the detailed analysis. We do not state all the assumptions, the results nor the proofs. We deliberately trivialize and simplify the models so as to make the discussions easier to follow. We aim to bring together separate but inherently related research in inventory literature. By doing so, we hope to highlight potential research opportunities that lie on the boundaries. We focus primarily on the author's previous work. The chapter does not aim to provide a review of the rich volume of publications. For that purpose, where possible, we refer the reader to comprehensive reviews.

The rest of the chapter is organized as follows. In § 2, we provide some examples of how managers can use information to better control inventory. In § 3, we consider the interaction between multiple inventory control managers and the economics of contracting. In § 4, we provide a discussion on large-scale inventory systems and rationality. In § 5, we provide some concluding thoughts and possible future research directions.

2   Information in Centralized Inventory Management

We will first discuss the use of information in centralized inventory management systems. An inventory management system is centralized when the system has access to credible information collected in a central location and managed by a single decision maker. Such a system is ideal; it does not have to coordinate disparate decisions and information. The manager needs to incorporate available information into the inventory control problem, identify the best replenishment policy and manage the system accordingly.

There are at least four reasons for studying centralized inventory systems. First, the results provide a benchmark against which decentralized inventory systems are measured. Second, the results

enable us to quantify and understand the role and value of information in inventory management. Third, small scale inventory systems are often centralized and are common in practice. Hence, it is necessary to know how to manage these systems. Industry has also learned the importance of centralized decision making such as the vendor managed inventory (VMI) initiatives. Fourth, the results also provide building blocks for large scale systems with decentralized operations.

To effectively manage inventory, a manager must have access to three fundamental sets of information (i) information about demand such as forecasts; (ii) information about assets such as the inventory available for sales, on order and where they are located; (iii) and information about replenishment lead times. In §2.1–2.4, we discuss single location inventory control problems, which are the minimal building blocks for multi location centralized inventory systems. We illustrate how the three fundamental sets of information are incorporated to develop effective production and inventory policies. We also show how managers can quantify the value of information by means of numerical computations. In §2.5, we provide a discussion on how these single-location inventory control models are used to study multi-location inventory systems.

2.1   Current Demand Information

We refer to demand information as *current* when the information is based on current data such as point of sales information and when it does not provide future information such as a promotion scheduled for next period, or advance order information. Here, we briefly review the classical single location inventory literature as a bridge to more recent work that incorporates the dynamic nature of demand information, such as forecast updates.

Early inventory models addressed the problem of minimizing ordering, holding, and backlogging costs for a single product at a single location over either a finite or an infinite horizon. Demand uncertainty is modeled as independent and identically distributed over time, i.e., demand D_t at each period t is an iid random variable. This modeling assumption uses current demand information.

In particular, the sequence of events for such a system is as follows. At the beginning of each period t, the manager reviews on-hand inventory I_t, any backorders B_t and the pipeline inventory. The manager decides whether or not to produce $z_t \geq 0$. She incurs a non-stationary production cost of $K_t\delta(z_t) + c_t(z_t)$, where $\delta(z) = 1$ if $z > 0$, K_t is the fixed production cost, and c_t is the variable production cost. The production initiated at period $t - L$ is added to the inventory, that is, L periods are required to complete the production. Demand D_t is observed. The demand for period t is satisfied through on-hand inventory; otherwise it is backordered. The manager incurs holding and penalty costs based on end-of-period net inventory.

Completing production takes L periods; hence, the manager needs to protect the system against the lead time demand $D_t^L = \sum_{s=t}^{t+L} D_s$. We let

$$x_t \quad : \quad \text{inventory position \underline{before} the production decision is made}$$
$$= \quad I_t + \sum_{s=t-L}^{t-1} z_s - B_t,$$
$$y_t \quad : \quad \text{inventory position \underline{after} the production decision is made}$$
$$= \quad x_t + z_t.$$

The expected holding and penalty costs charged to period t are given by $\tilde{G}_t(y_t) = \alpha^L E g_{t+L}(y_t - D_t^L)$,

where α is the discount factor and $g_t(x)$ is the single period holding and penalty cost based on inventory on hand at the end of period t. The expectation is with respect to the lead time demand D_t^L. It is assumed that g_t is convex and coercive for all t.[3] These properties are satisfied, for example, when a positive holding cost is charged per unit of inventory on hand and a positive penalty cost is charged per unit of backlog. The solution to the following dynamic programming recursion minimizes the cost of managing this single item, single location system for a finite horizon problem with $T - t$ periods remaining until termination.

$$J_t(x_t) = \min_{y_t \geq x_t} \{K_t \delta(y_t - x_t) + G_t(y_t) + \alpha E J_{t+1}(y_t - D_t)\},$$

where $J_{T+1}(\cdot) \equiv 0$ and $G_t(y_t) = (c_t - \alpha c_{t+1})y_t + \tilde{G}_t(y_t)$.[4]

Scarf (1959) characterizes the optimality of an (s, S) policy. Under this policy the manager orders up to S_t whenever the inventory position x_t falls below a critical level s_t. Veinott (1966) proves the optimality of (s, S) policies under different conditions. Infinite horizon results are due to Iglehart (1963). When the fixed cost of ordering is negligible, i.e., $K = 0$, an optimal policy is the base-stock policy with base-stock level S_t. Karlin (1960) and Veinott (1965) generalize the problem to account for seasonal variations in demand and non-stationary data and prove the optimality of period dependent base-stock policy. We refer the reader to Porteus (1990) for a review of classical inventory models.

Such policy parameters can often be obtained by a backward induction algorithm. A remarkable result that significantly reduces the computational burden is the optimality of a myopic policy that minimizes the current period inventory cost. Karlin (1960) and Veinott (1965) show that a myopic policy is optimal when the problem is stationary[5]; demand is stochastically increasing over time; or the myopic base-stock levels are increasing[6]. Morton and Pentico (1995) provide empirical evidence of how a myopic policy performs under various non-stationary environments. They also propose close-to-optimal, near-myopic policies. Iida (2001) also shows that myopic policies are effective when data changes "slowly".

Noticing that historical demand information might be used to understand uncertain customer demand, several authors incorporated demand history into inventory control problems. Three groups of work capture this idea. The first group uses Bayesian models. Under these models Bayes' rule defines a procedure to update the distribution of demand as new information becomes available. To the best of our knowledge, Dvoretzky, Keifer and Wolfowitz (1952) were the first to use this approach. Scarf (1960), Azoury and Miller (1984), and Azoury (1985) extended this approach. The second group, Johnson and Thompson (1975), Miller (1989), and Lovejoy (1990), realized that the demand over consecutive periods might be correlated and used time series models to subsume demand dynamics. The third group incorporates Markov modulated demand to the above inventory control problem (see, for example, Song and Zipkin 1993, Beyer and Sethi 1997, Abhyankar and Graves 2001 and Atali and Özer 2005).

---

[3]A function $g : \mathcal{R} \rightarrow \mathcal{R}$ is coercive if $\lim_{|x| \rightarrow \infty} g(x) = \infty$.

[4]It is often assumed that leftover inventory at the end of the planning horizon T is salvaged for c_{T+1} per item. Veinott (1965) shows that the inventory control problem with linear salvage value can be converted into an equivalent problem with zero salvage. Here we report the result of this conversion.

[5]An inventory problem is said to be stationary if the cost and demand distributions are time invariant.

[6]We use the terms increasing and decreasing in the weak sense. Increasing means nondecreasing.

2.2 Advance Demand Information

Most businesses rely heavily on demand forecasts for production and inventory planning. Demand over time can be highly correlated. Forecasting methods can help identify such patterns. A group of scholars have incorporated the dynamic nature of forecast revisions into inventory control problems. Papers in this group include those of Hausman (1969), Graves et al. (1986), Heath and Jackson (1994), Güllü (1996) and Toktay and Wein (2001). All of these works show that incorporating demand updates to control problems reduces the cost of managing inventories by proposing control methods that are responsive to forecast information.

Recent advances in information technology have enabled managers to be more proactive and obtain *advance* demand information in addition to improving demand forecast. Different customers have varying willingness to wait for the orders they placed. A good example of this concept is Dell's online Intelligent Fulfillment initiative, which allows four different levels of response time to customer orders: (1) standard (conventional or 5 day promised order lead time) (2) value delivery (slower but lower shipping cost); (3) premium delivery (same day delivery); and (4) precision delivery (specific date). A portfolio of online customers with differing response time preferences gives rise to advance demand information (ADI). Comparing inventory models with and without ADI, a manager can quantify the value of demand information contained in ADI (Özer 2003).

Several plausible strategies can be used to obtain advance demand information. When people order a customized product, they expect to wait for the product to be customized to their request. This can be called a built-in ADI. Alternatively, a discount could be offered for early orders to segment the customer based on their willingness to wait. If pricing is not an option, special service incentives could be offered for early orders. For example, a major truck manufacturer in North America provides free maintenance (up to ten years) for third party logistic providers (such as UPS) who purchase trucks a few years in advance of using them. Essentially, we are seeking those customers who have a high sensitivity to customization, price or service, and who also have a lower sensitivity to lead time or waiting time. These are denoted by "A" in Figure 1. They are the possible source of ADI.
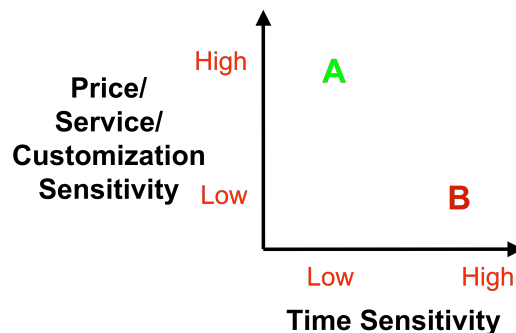

Figure 1: Source of ADI

Designing effective strategies to collect this information requires one to quantify the benefit of ADI. To do so, Gallego and Özer (2001), Özer (2003) and Özer and Wei (2004) show how to use this information optimally. In particular, they incorporate advance demand information into

periodic-review inventory control problems.

ADI is obtained when a customer places an order in any period t for delivery in a future period $s \in \{t+1, \ldots, t+N\}$. From the perspective of the production manager, the demand stream during period t is a vector:

$$D_t = (D_{t,t}, \ldots, D_{t,t+N}),$$

where $D_{t,s}$ represents the nonnegative demand for period s placed during period t and N is the length of the *information horizon*. Note that when $N = 0$, the problem reduces to the inventory problem with current demand information. This is a random vector and its uncertainty is resolved at the end of period t. Under this demand model, at the beginning of each period t, demand for a future period $s > t$ can be decomposed into two parts as illustrated in Figure 2: the observed part $O_{t,s} \equiv \sum_{r=s-N}^{t-1} D_{r,s}$ and the unobserved part $U_{t,s} \equiv \sum_{r=t}^{s} D_{r,s}$.
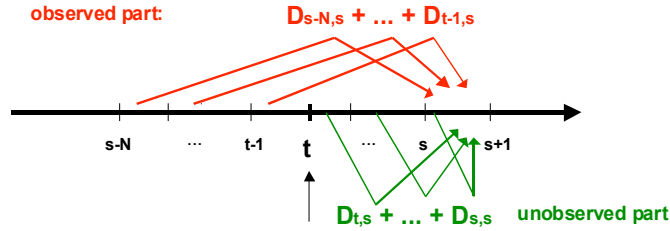

Figure 2: Observed and Unobserved Part of the Demand

The sequence of events is similar to the one described in the previous section. Completing production takes L periods; hence, the manager should protect the system against the lead time demand. Because of advance demand information, the manager knows part of the lead time demand, that is, $\sum_{s=t}^{t+L} O_{t,s}$. The expected cost charged to period t is based on the net inventory at the end of period $t + L$. Let

$$
\begin{aligned}
x_t^a \quad &: \quad \textit{modified} \text{ inventory position } \underline{\text{before}} \text{ the production decision is made} \\
&= \quad x_t - \sum_{s=t}^{t+L} O_{t,s}, \\
y_t^a \quad &: \quad \textit{modified} \text{ inventory position } \underline{\text{after}} \text{ the production decision is made} \\
&= \quad x_t^a + z_t.
\end{aligned}
$$

Notice that these variables subtract the observed part of the lead time demand, hence the name *modified*. In addition to x_t^a, the manager also keeps track of observations beyond the lead time, $O_t = (O_{t,t+L+1}, \ldots, O_{t,t+N-1})$. At the end of the period t, we update the state space by

$$x_{t+1}^a = y_t^a - D_{t,t} - \sum_{s=t+1}^{t+L+1} D_{t,s} - O_{t,t+L+1}, \tag{1}$$

$$O_{t+1,s} = O_{t,s} + D_{t,s}. \tag{2}$$

The expected holding and penalty cost charged to period t is given by $\tilde{G}_t(y_t) = \alpha^L E g_{t+L}(y_t - \sum_{s=t}^{t+L} U_{t,s})$. The solution to the following dynamic programming recursion minimizes the cost of managing this system for a finite horizon problem with $T - t$ periods remaining to the termination.

$$J_t(x_t, O_t) = \min_{y_t \geq x_t} \{K_t \delta(y_t - x_t) + G_t(y_t) + \alpha E J_{t+1}(x_{t+1}, O_{t+1})\}, \tag{3}$$

7

where $J_{T+1}(\cdot, \cdot) \equiv 0$ and $G_j(y_j) = (c_j - \alpha c_{j+1})y_j + \tilde{G}_j(y_j)$.

Gallego and Özer (2001) characterize the optimality of (i) a state-dependent (s, S) policy for an inventory system with positive fixed (set-up) costs and (ii) a state-dependent base stock policy for an inventory system without set-up costs both for finite and infinite horizon problems. The policy parameters depend on customer commitments made beyond the production leadtime. For example, if the production lead time is four periods, optimal policy parameters depend on the total customer commitments made today for delivery after four periods. Under this policy the manager produces up to S whenever the *modified* inventory position x_t^a drops to or below s. Gallego and Özer provide monotonicity results and characterize conditions when myopic policies are optimal. They also determine conditions under which ADI has no operational value. Through numerical studies and by comparing models with and without ADI, the authors quantify the benefit of inducing and obtaining ADI.

We note that incorporating advance demand information not only yields better practices through reducing inventories, but also enables companies to have control policies that are more responsive to changes in demand patterns. This information allows a shift from make-to-stock to make-to-order production. There is a growing body of research that shows how ADI can be used to improve costs in a capacity constrained system, continous review problems, or multi-echelon structures (Hariharan and Zipkin 1995, Schwartz et al. 1998, Gallego and Özer 2002, Karaesmen et al. 2002, Zhu and Thonemann 2003, Özer 2003, Özer and Wei 2004, Hu et al. 2004, Benjafaar et al. 2005, Marklund 2006 and Gayon et al. 2007). These models can be used to quantify the value of advance demand information in various settings. Being able to quantify its value, an inventory manager can decide how to optimally acquire advance demand information through pricing and advance sales and how to use this information in, for example, capacity decisions (Boyaci and Özer 2004). Such research also bridges the revenue management literature with the capacity management literature.

ADI and Capacity Management: We discuss how the results from the ADI literature were used to quantify the value of capacity and advance demand information for a global telecommunications equipment manufacturer. During the last quarter of 2002, this equipment manufacturer explored the strategy of advance selling to improve long-range forecasting for planning the capacity of a new factory. Accordingly, before securing the capacity the firm considered preselling wireless base-stations to its regional cellular phone operators.

The traditional view of capacity planning is that capacity is fixed, and lead times will vary to compensate for surges and gaps in orders. A different viewpoint is that one can fix and guarantee a lead-time; this requires the ability to flex capacity as needed. Figure 3 summarizes these two approaches. Suppose we had several types of customers as in Figure 4, where each class had different lead time requirements. Then we could guarantee lead times by customer segment, and implement this through careful scheduling of the facility. This strategy enables the firm to obtain advance demand information which can be used for better inventory and capacity planning.

The next issue is the management of the production system given the available capacity Q and the advance demand information. In order to minimize the cost of managing this production system, the manager maintains a safety stock. Recall that the manager would also like to satisfy some customers who have short lead times (even shorter than the production lead time) in addition to those who book well in advance. Hence one needs to maintain a safety stock. But what is the optimal level of inventory? More inventory means more money tied up, while less inventory may
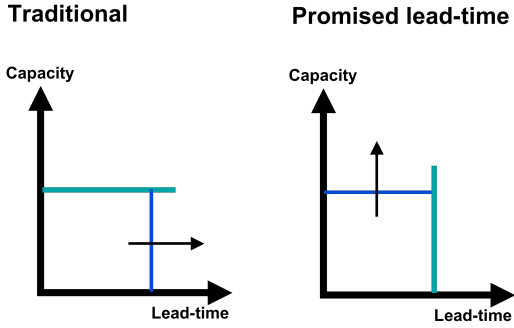
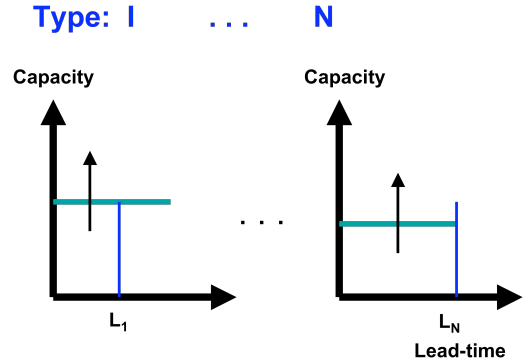Figure 3: Resource Planning: Always Having Capacity

Figure 4: Resource Planning: Fixed Capacity with Advance Demand Information

result in loss of customers and loss of goodwill. The solution to a dynamic program similar to the one in Equation (3) (but significantly more difficult to analyze) provides the best level of safety stock that minimizes the cost of underage and overage for a given planning horizon (see Özer and Wei 2004 for details).

Through an extensive numerical and simulation studies one can quantify the benefits gained through ADI for a capacity constrained system. In Figure 5 we provide one such example. The x axis shows the available capacity for production. The y axis shows the total inventory management cost (which is rescaled so that the cost of managing a system with infinite capacity and 100% ADI is zero). The three curves illustrate different levels of ADI. Curve A is the base case where all customers demand the product as soon as they place an order, whereas curve C has some customers who place their orders well in advance, and curve B is in between.
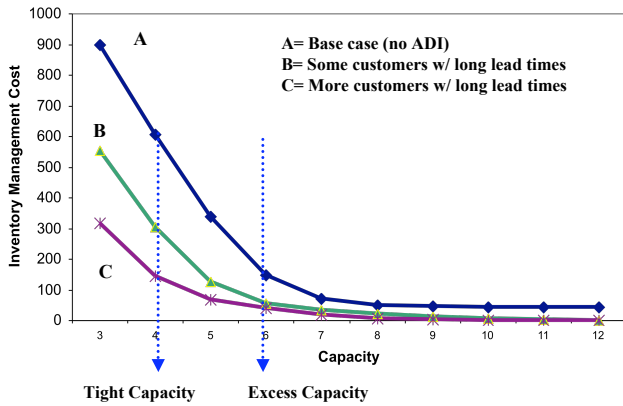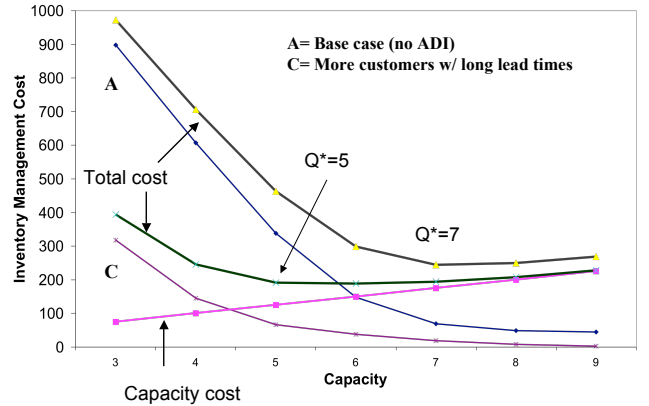

Figure 5: ADI versus Capacity

Figure 6: Optimal Capacity Size

We observe from each of these curves that additional capacity has diminishing returns. This suggests that there is an optimal level of capacity beyond which increasing capacity has limited operational value in managing this system. We also observe the reduction in the inventory man-

agement cost as a function of ADI and capacity. The vertical difference between these curves (for example between curves A and C) depicts the reduction in inventory costs due to employing advance demand information. Note that this reduction is more valuable when the firm is working under tight capacity.

Consider a capacity expansion (contraction) problem with capacity increment ΔQ. The expansion cost $C(\Delta Q)$ may take several forms, such as linear, power or step cost function (Luss 1982). If capacity expansion is a one time decision, then the manufacturer's problem is to solve $\min_{\Delta Q}\{C(\Delta Q) + J_t(x, O|Q + \Delta Q)\}$. Figure 6 provides an example of this problem when $C(\Delta Q) = 100 * \Delta Q$ under two advance demand information scenarios. For this particular example, convincing customers to place orders in advance reduces the optimal capacity expansion decision from $Q^* = 7$ to 5 units. This is another example illustrating how advance demand information can be a substitute for capacity (Özer and Wei 2004).

2.3 Imperfect Asset Information

Since the early 80's the availability of cheaper and faster computation enabled companies to automate their inventory management processes and to use inventory management softwares. Automatic replenishment systems track the number of products in stock and place replenishment orders based on the control policies set by the underlying software. A crucial assumption used by these inventory management systems is that inventory record and actual on-hand inventory are identical.

Similarly, the standard inventory control literature has never differentiated between inventory *record* and *actual* inventory. The two have always been considered to be the same. In the previous sections we assumed that the manager knows the exact value of, for example, the inventory position. The implicit assumption was that all demand sources are visible. However, recent surveys and empirical work have shown that unaccounted inventory due to, for example, theft or misplacement, can lead to a significant discrepancy between inventory records and actual inventory (as documented by empirical studies such as Rinehart 1960, Raman et al. 2001a,b, ECR Europe 2003). As a result, stock-outs are widespread at retailers and distributors (Alexander et al. 2002).

Early modeling approach for inventory control under imperfect asset information is due to Iglehart and Morey (1972). They study the impact of transaction errors only and do not consider misplacement or shrinkage. They also decompose the error management problem from inventory management. In particular, they establish the approximate buffer stock required to hedge against transaction errors independent of the buffer stock necessary to hedge against paying customer demand. Kang and Gershwin (2005) consider discrepancy due to shrinkage and its impact on inventory management through a simulation study. We refer the reader to Lee and Özer (2007) for a detailed discussion of inaccuracy problems. Through model based analysis, the authors quantify the benefit of a tracking technology known as RFID in supply chain management.

In a recent paper, Atali, Lee and Özer (2004, 2006) characterize three different kinds of demand streams that result in inventory discrepancy. Some demand streams result in permanent inventory shrinkage (such as theft and damage). They refer to this stream as *shrinkage*. Some demand streams are temporary and can be recovered by physical inventory audit and returned to inventory (such as misplacement). They refer to this demand stream as *misplacement*. The final type of demand stream (such as scanning error) affects *only* the inventory record and leaves actual inventory unchanged.

They refer to this stream as *transaction errors*. It is necessary to characterize these sources separately because each of these sources affects the system in a unique way. For example, misplaced items can be returned back to inventory after an inventory audit, whereas stolen items cannot.

There are four ways to manage an inventory system that faces an inventory discrepancy problem, summarized in Figure 7 (Atali, Lee and Özer 2004, 2006). The first way is to ignore the discrepancy problem and use only the point of sales data information to drive the replenishment process. The second way is to use the statistics about unobservable demand sources in driving the replenishment process, for example, by carrying additional buffer stock to hedge against transaction errors. The third way is to invest in a technology such as RFID that enables complete visibility of inventory movement and use the actual information (instead of the statistics) to drive the replenishment process. The fourth way is to go one step further and use the visibility to prevent or reduce unobservable demand sources, for example, by locating and reshelving misplaced items as soon as a customer misplaces the item).

<div align="center">

Visibility

	Without	*With*
Passive	**Base Case** POS-driven replenishment	**RFID I Case** Replenishment based on accurate inventory
Active	**Smart Case** Adjustment based on statistical parameters	**RFID II Case** Actively correct misplaced items

Control

</div>

Figure 7: Inventory Management Cases under Imperfect Asset Information

Here we discuss the formulation of the third case (RFID I case in Figure 7) and refer the reader to Atali, Lee and Özer (2004, 2006) for a detailed treatment of the other cases. The sequence of events for this case is as follows. At the beginning of period t, the inventory manager reviews the state of the system and decides to order $z_t \geq 0$ units from an outside supplier with ample supply. The replenishment lead time is assumed to be zero. The cost of ordering is c_t per unit. There is no fixed cost for placing an order. Purchasing customer D_t^p, misplacement D_t^m, shrinkage D_t^s, transaction errors D_t^τ arrive in any sequence. Note that the realizations of these error sources are observed because we are considering the third way to manage the inventory system. At the end of the period, the manager incurs a linear holding cost h_t and a linear lost-sales cost p_t based on the end of period physical on-hand inventory. Holding cost is incurred for the misplaced items even though they are not available for sales. No lost-sales cost is incurred for unmet demand from nonpaying customers. If the period is a counting (audit) period, an inventory audit is conducted at the end of that period. The inventory record is reconciled: error is corrected, and all misplaced items are returned to inventory. Otherwise, errors continue to accumulate. The planning horizon is a multiple of counting cycle length, that is, $T \in \{N, 2N, 3N, ...\}$. At the end of the planning horizon T, the inventory left over is sold for a linear salvage value of c_{T+1}.

At the beginning of period t, the manager knows the inventory *record* x_t^r, the accumulated error terms e_t^s, e_t^m, e_t^τ and the number of periods elapsed since the last inventory count, i_t. The state space

of such as system can be summarized by (x_t, e_t^m, i_t), where

$$x_t = x_t^r - e_t^m - e_t^s - e_t^\tau$$

is the *sales-available* on-hand inventory, and $i_t \in \{0, 1, ..., N-1\}$. The state of the system evolves according to the following equations.

$$x_{t+1} = \begin{cases} [y_t - D_t]^+, & \text{if } i_t \neq N-1 \\ [y_t - D_t]^+ + e_t^m + m_t, & \text{if } i_t = N-1 \end{cases} \tag{4}$$

$$e_{t+1}^m = \begin{cases} e_t^m + m_t, & \text{if } i_t \neq N-1 \\ 0, & \text{if } i_t = N-1 \end{cases} \tag{5}$$

$$i_{t+1} = (i_t + 1) \bmod N, \tag{6}$$

where $y_t = x_t + z_t$ and $D_t = D_t^p + D_t^s + D_t^m$ and m_t is the realized misplacement. The single period expected holding and penalty cost charged to period t is based on *sales-available* on-hand inventory and the accumulated misplacement.

$$\tilde{G}_t(y_t, e_t^m) = h_t E_{D_t, m_t}([y_t - D_t]^+ + e_t^m + m_t) + p_t E_{D_t^p, a_t}(D_t^p - a_t). \tag{7}$$

Transaction errors are random observation disturbances and they have no direct impact on the *sales-available* on-hand inventory x_t.

With perfect visibility, the manager optimizes the stock levels in full awareness of the inventory errors that take place during period t. Let J_t^v be the cost of managing this system for a finite horizon with $T - t$ periods remaining to the end of the planning horizon. The optimal replenishment policy would be to select the value of y_t that minimizes the following dynamic programming algorithm.

$$J_t^v(x_t, e_t^m, i_t) = \min_{y_t \geq x_t} \{G_t(y_t, e_t^m) + \alpha E J_{t+1}^v(x_{t+1}, e_{t+1}^m, i_{t+1})\}, \tag{8}$$

where $J_{T+1}^v(x_{T+1}, ., .) = 0$ and $G_t(y_t, e_t^m) = c_t y_t - \alpha c_{t+1} E x_{t+1} + \tilde{G}_t(y_t, e_t^m)$. The leftovers at the end of the planning horizon $T + 1$ are salvaged for a linear price.

To calculate the aforementioned expectations in the dynamic programming algorithm, one needs to obtain the distribution of sales a_t and misplacement m_t during any period t. However, the realization of these variables and their distribution depend on the sales-available on-hand inventory x_t and the order in which misplacement, shrinkage and paying customer demands arrive.

Consider a modified model in which the paying customer demand always arrives first, demand for shrinkage arrives next and demand for misplacement arrives last. With this sequence, sales during any period are maximized while misplacement is minimized. The transaction error can arrive in any where in the sequence because it does not affect the physical inventory. Given this sequence, the sales and the misplacement during period t are

$$a_t = \min\{D_t^p, y_t\}, \tag{9}$$

$$m_t = \min\{D_t^m, [y_t - D_t^p - D_t^s]^+\}. \tag{10}$$

The state of the system evolves according to the Equations in (4-6), but with m_t replaced by its new definition above. Similarly, the single period cost function is the same as in Equation (7) but with a_t and m_t replaced by their respective definitions. Using similar demand prioritization ideas, one

12

can construct bounds and effective solutions for the above dynamic programming problem. They enable effective inventory control methods when the manager uses RFID or a similar technology that provides complete visibility of inventory movement in the store.

Atali, Lee and Özer (2004) characterize efficient replenishment policies for all four cases in Figure 7. Using these models and comparing the resulting cost of each scenario, they quantify the *true* value of visibility provided by RFID. Consider, for example, the value of visibility. When the system does not use a technology such as RFID, the manager can use; either the *informed* policy that corresponds to the smart case, or an *ignorant* policy that corresponds to the base case in Figure 7. Recall that in the base case the replenishment policy is obtained without consideration of the discrepancy problem whereas informed policy uses some statistics about discrepancy. The *true* value of visibility is given by the cost difference between the informed policy and the policy that uses visibility, i.e., RFID I or II.
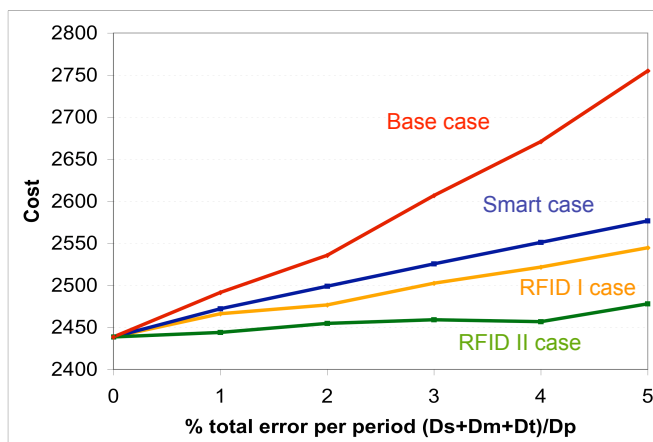

Figure 8: Value of Visibility and Active Control as a Function of Total Error Source

Figure 8 compares the resulting cost for a problem instance as a function of total error with respect to paying customer demand, i.e., total average error divided by average paying customer demand. The lowest curve is the cost of following an effective replenishment policy when the manager has complete visibility of inventory and follows an active control strategy. This figure illustrates that by using an informed policy to compensate for the discrepancy problem, the manager can reduce costs significantly. The value of visibility also increases with the total percentage errors. For a particular example, when compared to the ignorant policy (base case), the visibility enabled system reduces cost by 9.1% and increases sales by 1.8%. However, when compared to the smart policy, the cost is reduced by 3.1% and the sales is increased by 0.1%. For this system, assuming that visibility also enables one to reduce the shrinkage rate by 50%, the manager can save (the difference between the visibility enabled systems with different shrinkage rates) an additional 2.6%, and increase sales by 0.1%, both of which can be interpreted as the value of prevention due to visibility brought by a technology such as RFID.

Recently, Atali, Lee, and Özer (2006) model demand streams using a random disaggregation model. In particular, let D_t denote the random customer demand during period t. An arriving customer buys the product with probability θ_p; misplaces the item with probability θ_m; or damages/steals the item with probability θ_s such that $\theta_p + \theta_m + \theta_s = 1$ for all t. Both demand modeling approaches have their own appeal. Random disaggregation approach simplifies the previous analysis.

13

In particular, one does not need to construct bounds through demand prioritization. Calibrating the model and fitting data is relatively simpler as well. However, the previous approach allows for independent demand streams for paying and non-paying customers.

2.4 Lead Time Information

When lead times are uncertain, information on the location of the supply plays a critical role. Classical inventory models assume that lead time for an order is independently drawn from a given distribution. Kaplan (1970) and Ehrdhardt (1984) discuss two assumptions that allow optimal control policies analogous to classical results with deterministic lead times. These assumptions are (i) deliveries of orders cannot cross in time and (ii) the delivery leadtime is independent of the number and size of outstanding orders. Anupindi, Morton and Pentico (1996) provide close-to-optimal, near-myopic heuristics to solve stochastic lead time inventory control problems. Janakiraman and Roundy (2004) provide some convexity results that enables the use of search procedures to determine optimal base-stock levels (see Chapter 7 of Zipkin 2000 for a comprehensive review of the stochastic lead time inventory control problems).

Song and Zipkin (1996) model the supply process as a Markov chain. The lead time L_t for an order in period t is a Markov chain with a finite state space. The transition matrix is such that orders are received in the order they were shipped (i.e., order cross over is not allowed). They assume that the inventory manager has visibility of the supply process (i.e., the Markov state) at the beginning of each period. She uses this information to revise the inventory ordering decision. The authors show that a state-dependent replenishment policy is optimal. Chen and Yu (2005) consider the problem in which the manager does not know the status of the system, but knows that the lead time is generated by a Markov process. Chen and Yu show that the value of lead time information is small for slow moving items. However, it can be as high as 40% for fast moving items. To demonstrate this, they numerically compare the model in which lead time is observable to that of Song and Zipkin.

The conventional modeling approaches for stochastic lead times generally assume that the statistical information essentially boils down to the mean and standard deviation of the lead time, and the safety stock takes into consideration such statistics. Recent information technologies, however, enable a manager to collect some advanced knowledge about the lead time as the product progresses over intermediate points, known in logistics as "choke points". Through tracking technologies and well connected computer networks, a manager can follow the progress of a supply before it reaches the store. Gaukler, Özer and Hausman (2004) quantify the benefit of this supply progress information. They propose and evaluate a replenishment policy that uses order progress information for emergency ordering together with the (Q,R) policy. In particular, the manager places a *regular* replenishment order of size Q when the inventory position drops to the reorder level R. They model the sojourn time for a regular order to move from one choke point to the next with a general non-identical distribution and provide additional results for the exponential distribution case. In addition, the manager also has the option to place an *emergency* order at a cost premium $K(l)$ of size αQ which arrives after a deterministic lead time l. They characterize the optimality of a state-dependent threshold policy for releasing an emergency order. In particular, the retailer monitors the outstanding regular orders location in the supply system, that is, the last choke point where the regular order was registered (of course, if a regular order is outstanding). If the inventory position

is less than a state-dependent threshold $\bar{y}$, the retailer places the *emergency* order. The threshold depends on where the regular order was registered last. Through a numerical study, the authors report overall cost savings ranging from 2.8-5.5% due to supply progress information. They show that the emergency ordering option eliminate up to 99% of the cost due to backlogging a customer. See also Moinzadeh and Schmidt (1991) and Moinzadeh and Aggarwal (1997) for the use of emergency ordering in single and multi-echelon inventory systems.

2.5 Multi-Location Inventory Systems

Clark and Scarf (1960) initiated the study of multi-echelon inventory systems. They show that a serial system can be optimally decomposed into *single* location problems and characterize the optimality of echelon base-stock policies. Under this policy a central inventory manager observes the echelon inventory position of each location and places an order from the outside supplier if the first echelon's inventory position is below its base-stock level. The manager also pushes inventory (as much as possible) to the downstream location j from its immediate predecessor if location j's echelon inventory position is less than its echelon base-stock level. Federgruen and Zipkin (1984), extend the results for stationary infinite horizon problems. Chen and Zheng (1994) establish lower bounds on the average cost and construct feasible policies that achieve these bounds. Unlike the single location inventory control literature, the multi-echelon in series literature lacks models that incorporate historical demand information. Chen and Song (2001) write the only paper to study the serial system with a non-stationary demand process, which is modulated by a finite-state, exogenous Markov chain. Graves et al. (1998) provide heuristic allocation of inventories across a serial system that obtains a forecast over a finite horizon. Gallego and Özer (2004) incorporate advance demand information into multiechelon, inventory systems in series and prove the optimality of state-dependent, echelon base-stock policies for finite and infinite horizon problems. The authors show that under certain conditions a *myopic* policy is optimal for a finite horizon multi-echelon inventory problem in series with and without advance demand information. This result significantly reduces the computational burden required to solve such serial systems. These systems are also fundamental to the study of more general structures. For example, Rosling (1989) shows that under mild conditions *assembly* systems can be treated as serial systems.

One of the most common multi-echelon structures in practice are the distribution systems, which are also known as one-warehouse-multi-retailer systems. Products enter the system from an outside supplier to the warehouse, which in turn replenishes various retailers. Stochastic demand is satisfied as much as possible through on hand inventory at the retailers. Clark and Scarf (1960) show that optimal control policies, if they exist, would be very complex for distribution systems. Since then the research on distribution systems has shifted towards identification of close-to-optimal heuristics and evaluation of a plausible class of policies. There are two approaches to solve this problem: approximation by relaxation as in Federgruen and Zipkin (1984a), Aviv and Federgruen (2001a), Özer (2003) and approximation by restriction as in Eppen and Schrage (1981), Federgruen and Zipkin (1984b) and Özer (2003). The first approach considers relaxing a constraint set to obtain a simpler problem with lower-dimensional state space. It develops a heuristic based on this lower bound problem to solve the original problem. The second approach restricts the policy space to a class of policies and optimizes over this class under additional assumptions. The restriction approach, unlike the relaxation approach, does not guarantee any bound on the optimal solution.

Other researchers that use approaches that do not guarantee any bounds include Diks and de Kök (1998). Comprehensive earlier reviews can be found in Axsäter (1993) for continuous review (also known as pull) and Federgruen (1993) for periodic review (also known as push) inventory systems.

In distribution systems, the "warehouse" may serve as the coordination center. It may also help negotiate lower procurement prices. Eppen and Schrage (1981) illustrate that the warehouse also serves an important enabler for *statistical economies of scale*, commonly known as risk pooling, that is, the portfolio effect of coordinating inventory decisions and holding inventory at the distribution center rather than at the retailers. Aviv and Federgruen (2001) incorporate a Bayesian framework into the demand process and introduce the concept of *learning effect* to the benefit of having a central distribution center. The ability to obtain information about the demand during the first periods enables updating the demand process, resulting in improved allocation to retailers. Özer (2003) incorporates advance demand information structure obtained from customers through each retailer. The author establishes a close-to-optimal state dependent replenishment and allocation policy that responds to the changes in customer demand. The author also provides a closed-form solution to approximate the system-wide inventory level. Using such explicit solutions, the model and the heuristic, he quantifies, for example, the benefit of advance demand information and its impact on allocation decisions and the joint role of risk pooling and advance demand information. For a review of these approaches we also refer the reader to Özer (2003).

The distribution system described here can also be interpreted as a multi-item production system with a common intermediate product. In this interpretation, the warehouse represents the differentiation point. During the first phase of the production a common batch is produced. At the end of this phase, the manager must decide on how much of each differentiated item to produce from the batch of the common intermediate product. This interpretation forms the basis of postponement strategies; see the papers by Lee, Padmanadbhan and Whang (1993) and Lee and Tang (1997). We conclude this section by noting that at the heart of all complex inventory systems lies the single location (stage, product, item) model that we addressed in the previous subsections.

3   Information in Decentralized Inventory Management

Global operations involve several locations managed by several inventory managers. The decisions and information are often decentralized. Many experts have heralded advances in information technology and Internet infrastructure, both of which enable better visibility and information sharing, as the key to effective management of inventory. Suppliers and manufacturers can share private information regarding, for example, costs or forecasts, but will they want to? Firms may be reluctant to collect, process and share information because of conflicting incentives. Aligning incentives improves firms profits and sustains the use of information technology.

Inventory managers can use formal contracts to align incentives and induce information sharing. There are two forms of information asymmetry. The informed party may withhold information to gain strategic advantage. In such cases, the uninformed partner can propose a menu of contracts to extract this information; this interaction is known as *adverse selection* or *screening*. Alternatively, the informed party may signal his information to gain cooperation. However, he needs to signal private information in a *credible* way; this interaction is known as *signaling* game. Another form of information asymmetry arises as moral hazard where one partner influences system profit through an

action or choice not observable to the other. The non-acting partner designs a contract to maximize his own profit (Fudenberg and Tirole 1991 and Salanie 1997). This section provides examples of such interactions in inventory management.

We attribute incentive problems in supply chains to lack of credible information sharing and three major risk imbalances: capacity risk, inventory risk and quality risk (Özer 2004). Because of lack of credible information sharing, the adverse effects of inventory and quality risks are more severe for a decentralized supply chain than for a vertically integrated supply chain. Here, we discuss some recent and ongoing research in designing contracts to eliminate or mitigate these adverse effects. We typify a two level supply chain by referring to an upstream member as the supplier and the downstream as the manufacturer.

3.1 Capacity Risk

Here we summarize results from Özer and Wei (2006). Forecasting demand is inherently difficult due to short product life cycles and long production lead time. Hence, supply chains face the risk of either excess capacity due to low demand realization (downside risk) or lack of product availability due to high demand realization (upside risk). Consider a manufacturer who builds to order and requires the supplier to deliver just in time. To deliver on time, the supplier secures component capacity or inventory in advance of a manufacturer order. If consumer demand turns out to be high, both the supplier and the manufacturer face upside capacity risk. However, if consumer demand turns out to be low, only the supplier faces downside capacity risk. Lack of proper risk sharing exacerbates the cost of capacity risk.

Double Marginalization:

The severity of capacity risk for each party depends on the contractual agreements. Under a wholesale price contract, for example, the manufacturer pays a wholesale price w to the supplier for each unit ordered. The supplier decides on the component capacity K to maximize his profit prior to observing demand. Let c_k be the unit cost of capacity. This cost could also represent an equivalent annual cost of capacity. Demand D is realized and the manufacturer places an order. The supplier fills the order as much as possible, at a unit cost c; that is, he delivers $\min(D, K)$. The manufacturer receives the order and sells at a fixed price $r > 0$.[7] Suppose unmet demand is lost without additional stock out penalty, and unsold inventory has zero salvage value without loss of generality.

Note that demand D is uncertain at the time when the supplier builds capacity. Suppose the demand forecast is such that $D = \mu + \epsilon$, where μ is the mean, which is a positive constant, and ϵ is a zero mean random variable with a cdf $G(\cdot)$, which represents the market or forecast uncertainty. Such information can be constructed by using information obtained, for example, through a third-party market research firm (such as Dataquest services of Gartner group). For a given capacity K, the manufacturer's and the supplier's expected profit before demand is realized are given by

$$\Pi^m(K) = (r - w)E\min(D, K), \tag{11}$$

$$\Pi^s(K) = (w - c)E\min(D, K) - c_k K. \tag{12}$$

---

[7]The manufacturer may carry out some value added operations that cost, say m per unit. She sells at a fixed unit price $r' > 0$. So her effective sales price is $r = r' - m$. Hence, without loss of generality, we assume $m = 0$. Of course, the story would be different if the manufacturer were building to stock as we will discuss in §3.2.

The supplier maximizes his profit in (12) by setting capacity to

$$K^w = \mu + G^{-1}(\frac{w - c - c_k}{w - c}).$$

Next consider the centrally integrated supply chain in which a single firm owns the manufacturer and the supplier. This centralized firm's expected profit and it's optimal capacity would be

$$\Pi^{cs}(K) = (r - c)E \min(D, K) - c_k K, \tag{13}$$

$$K^{cs} = \mu + G^{-1}(\frac{r - c - c_k}{w - c}). \tag{14}$$

Note from (12) and (13) that the supplier's marginal profit is less than the vertically integrated supply chain's marginal profit. This difference is due to double marginalization. The supplier, therefore, secures less capacity than what would be optimal for a vertically integrated supply chain, that is $K^w \leq K^{cs}$. Note that $\Pi^{cs}(K^{cs}) \geq \Pi^m(K^w) + \Pi^s(K^w)$. Hence, both the manufacturer and the supplier are leaving money on the table due to decentralized operations. The magnitude of this inefficiency depends on the parameters.

The manufacturer may encourage the supplier to build more capacity by providing some protection against the downside risk, the risk of having excess capacity. Observe that the manufacturer's payoff (the realized profit) is always nonnegative, while the supplier faces the risk of a negative payoff. The manufacturer can share this risk by providing a payment in case of excess capacity after demand is realized. One such contract is the *payback* contract (w, τ), under which the manufacturer pays the supplier w per unit for its order and τ per unit for unused capacity $(K - D)^+$.

The manufacturer's and the supplier's expected profit functions for this case are

$$\Pi^m(K) = (r - w)E \min(D, K) - \tau E(K - D)^+,$$
$$\Pi^s(K) = (w - c)E \min(D, K) + \tau E(K - D)^+ - c_k K.$$

The supplier solves $\max_{K \geq 0} \Pi^s(K)$. The optimal capacity is $K^\tau \equiv \mu + G^{-1}(\frac{w - c - c_k}{w - c - \tau})$. To achieve channel coordination, we equate K^τ with K^{cs} and solve for τ, resulting in $\tau = \frac{(r - w)c_k}{r - c - c_k}$. Hence, the payback contract (w, τ) can coordinate the channel, that is the sum of the manufacturer's and supplier's profit under this contract is equal to the profit of the centralized firm. The supplier captures $\frac{w - c - c_k}{r - c - c_k} \times 100\%$ of the total profit; and the manufacturer captures $\frac{r - w}{r - c - c_k} \times 100\%$ of the total profit. Notice that arbitrary profit division among the parties is also achievable by changing the wholesale price w. Therefore, with the appropriate choice of (w, τ) the payback contract results in mutually beneficial terms, that is, the manufacturer's and the supplier's expected profits are at least as large as their profits under any wholesale price contract[8].

Asymmetric Forecast Information

Another issue in the aforementioned supply chain is the forecast sharing problem. The wholesale price is often set during the product design stage, which takes place long before the manufacturer ramps up the production. Component capacity or inventory commitments, however, are often made closer to production. Hence, the forecast sharing problem is often decoupled from product design and wholesale price negotiations. The manufacturer often has better forecast information than the

---

[8]The payback contract provides a reward mechanism that induces the supplier to secure more capacity. Another mechanism is to penalize the supplier for every unit of order that he is unable to satisfy due to capacity shortage.

supplier due to her proximity to consumers. Lee, Padmanadbhan and Whang (1997) provide four reasons, such as order batching, for why the downstream member distorts the demand forecast when sharing it with the upstream member. Özer and Wei (2006) show that another key reason for the bullwhip is the form of the contract.

Suppose that the aforementioned manufacturer has new forecast information before the supplier sets the capacity. Let ξ denote the manufacturer's *private* information about demand forecast. Suppose ξ is a deterministically known quantity to the manufacturer. The manufacturer's new demand forecast information is $D = \mu + \epsilon + \xi$. If the supplier has access to the manufacturer's private forecast information ξ, he maximizes (12) by setting the capacity to

$$K^w = \mu + \xi + G^{-1}\left(\frac{w - c - c_k}{w - c}\right). \tag{15}$$

However, ξ is known *only* to the manufacturer. Can the manufacturer share this forecast information *credibly*? The answer is no because the manufacturer has an incentive to inflate her report of ξ. This incentive arises because the manufacturer's profit in (11) is increasing in the supplier's capacity choice K and the suppliers optimal capacity K^w is increasing in the manufacturer's forecast information ξ. Hence, by sharing an inflated forecast the manufacturer can increase her expected profit. The supplier, therefore, would never consider the forecast information provided by the manufacturer to be *credible* regardless of the manufacturer's sincere effort to share her forecast information. Instead, the supplier would resort to his *prior* belief about the manufacturer's private forecast information. For example, the supplier may perceive ξ to be a zero mean random variable that takes values in $[\underline{\xi}, \bar{\xi}]$ with cdf $F(\cdot)$.

This concept is what is known as *asymmetric* forecast information. The supplier and the manufacturer have asymmetric information about ξ and hence the overall demand forecast. The manufacturer knows ξ deterministically, whereas the supplier has a prior *belief* about its possible value. Hence, the supplier's expected profit is

$$E_\xi \Pi^s(K, \xi) = (w - c)E\min(\mu + \xi + \epsilon, K) - c_k K, \tag{16}$$

where the uncertainty is due to <u>both</u> ξ and ϵ. The supplier maximizes (16) by setting capacity level

$$K^{wa} \equiv \mu + (F \circ G)^{-1}\left(\frac{w - c - c_k}{w - c}\right), \tag{17}$$

where $F \circ G$ is the distribution function of $\xi + \epsilon$.

Comparing the supplier's capacity decision when she has and does not have access to the manufacturer's forecast information reveals the source of inefficiency. The supplier's capacity choice without having access to ξ under asymmetric information K^{wa} is not a function of ξ. Without credible forecast information sharing, the supplier cannot adjust the capacity to account for the manufacturer's private forecast. The consequences of this inefficiency could be severe for *both* parties. When the manufacturer's private forecast is very high, both parties may lose sales, resulting in lower profits (as the Boeing case in Cole 1997). When the manufacturer's private forecast information is low, the supplier may suffer from excess capacity (as the Solectron case in Hibbard 2003). The remedy for *this* inefficiency is to induce *credible* information sharing.

Özer and Wei (2006) show that the supplier can hold the manufacturer accountable for her private forecast information by requiring a *monetary* commitment before securing component capacity. This

accountability can be achieved by designing a menu of prices for reserving capacity. The menu should be designed in a way that the supplier can *screen* the manufacturer's forecast information. To do so, the supplier offers this menu any time before setting the capacity.

The sequence of events is as follows. The supplier provides a *menu of contracts* $\{K(\xi), P(\xi)\}$ for all $\xi \in [\underline{\xi}, \bar{\xi}]$. Both capacity and corresponding payment are functions of private forecast information ξ. Here, the supplier's objective is to find the optimal menu that maximizes his profit. Given this menu, the manufacturer chooses a particular contract $(K(\hat{\xi}), P(\hat{\xi}))$ that maximizes her profit. By doing so, she announces her forecast information to be $\hat{\xi}$, which could differ from her true forecast information ξ. The supplier receives the payment $P(\hat{\xi})$ and builds capacity $K(\hat{\xi})$ at unit cost c_k. The manufacturer observes demand D and places an order. The supplier produces as much of the order as possible given the capacity constraint; that is, he delivers $\min(D, K(\hat{\xi}))$. The manufacturer receives the order and sells at unit price $r > 0$. Two decisions are the supplier's choice for the optimal *menu of contracts* that maximizes his profit; the manufacturers' choice from this menu is the optimal *contract* that maximizes her profit.

By choosing a contract, the manufacturer defines her profit, the supplier's profit and the total supply chain profit as

$$\Pi^m(K(\hat{\xi}), P(\hat{\xi}), \xi) = (r - w)E\min(\mu + \xi + \epsilon, K(\hat{\xi})) - P(\hat{\xi}), \tag{18}$$

$$\Pi^s(K(\hat{\xi}), P(\hat{\xi}), \xi) = (w - c)E\min(\mu + \xi + \epsilon, K(\hat{\xi})) + P(\hat{\xi}) - c_k K(\hat{\xi}). \tag{19}$$

The supplier's challenge is to elicit truthful information and to maximize his profit by choosing a menu of contracts while ensuring the manufacturer's participation. To identify an optimal menu of contracts, the supplier solves

$$\max_{K(\cdot), P(\cdot)} E\Pi^s(K(\xi), P(\xi), \xi) \tag{20}$$

$$s.t. \quad \textbf{IC:} \quad \Pi^m(K(\xi), P(\xi), \xi) \geq \Pi^m(K(\hat{\xi}), P(\hat{\xi}), \xi), \text{for all } \hat{\xi} \neq \xi$$

$$\textbf{PC:} \quad \Pi^m(K(\xi), P(\xi), \xi) \geq \pi^m_{min}, \text{for all } \xi \in [\underline{\xi}, \bar{\xi}].$$

The expectation in the supplier's objective is with respect to ξ. The first set of constraints is the incentive compatibility (IC) constraints. These constraints ensure that the manufacturer maximizes her profit only by truthfully revealing her forecast information. The second set of constraints is the participation constraints (PC). They ensure a minimum profit π^m_{min} to the manufacturer regardless of her forecast information. This minimum profit could be the manufacturer's profit from her outside option, or her profit under other contracts. Note that this optimization problem is a difficult one that involves optimization over functions. Through obtaining structural results, this problem can be converted to equivalent but gradually simpler formulations that are easier to deal with (see Özer and Wei 2006).

The authors provide closed-form solutions/formulas as a solution to the the optimization problem in (20). They also show that the optimal $P^{cr}(\xi)$ and $K^{cr}(\xi)$ are monotone in ξ. Hence, one can construct a function $P(K)$ by setting $P(K) = P^{cr}(\xi)$, if $K = K^{cr}(\xi)$. This function can be interpreted as a *capacity reservation* contract; i.e., pay $P(K)$ to reserve K units of capacity. Note that the optimal contract is independent of the manufacturer's forecast information. The supplier simply gives this contract as a menu of fees for the corresponding capacity level that the manufacturer

may reserve. Essentially the supplier delegates the capacity decision right to the manufacturer, who has superior forecast information.

The supplier can also hold the manufacturer accountable for her private forecast information by requiring a *quantity* commitment before the supplier secures component capacity. Özer and Wei structure the *advance purchase* contract under which the manufacturer pays the supplier w_a for each unit she orders before the supplier secures capacity; hence the name, advance purchase. This agreement provides an option to the manufacturer to place firm orders at an advance purchase price before the supplier secures capacity. The advance purchase could be costly to the manufacturer if the realized demand turns out to be smaller than the advance purchase quantity. Intuitively, this commitment prevents a manufacturer with a low forecast from communicating a high forecast. Özer and Wei (2006) show that the manufacturer can credibly *signal* her forecast through placing an advance purchase before the supplier decides the capacity. The authors also show that channel coordination is possible *even* under asymmetric forecast information by combining the advance purchase contract with an appropriate payback agreement. The formulation and analysis of the advance purchase contract leads to a *signaling* game, whereas the capacity reservation contract is a *screening* game. By comparing these models and analysis, the authors also show analytically when to use these contracts.

Which Contract Form to Adopt?

Özer and Wei (2006) identify two key drivers of the (supplier's, manufacturer's and supply chain's) expected profits under different contracts: the *risk adjusted profit margin* and the *degree of forecast information asymmetry.*

Recall that the supplier's profit margin is less than the integrated supply chain's profit margin per unit of capacity investment. Hence, the supplier builds less than the supply-chain-optimal capacity. Two factors determine the impact of this inefficiency on the supply chain: the market uncertainty modeled by ϵ, and the supplier's profit margin per unit *sold* or per unit of capacity built, that is $w - c - c_k$. Hence, this inefficiency can be measured by the *risk-adjusted profit margin* $\frac{(w-c-c_k)}{\sigma_\epsilon}$, that is, the supplier's profit margin per unit sold per unit of market uncertainty.

The severity of supply chain inefficiency also depends on how much the supplier knows about demand as compared to the manufacturer. This knowledge disparity is measured by the *degree of forecast information asymmetry.* Let σ_ξ and σ_ϵ be the standard deviations of $F(\cdot)$ and $G(\cdot)$, respectively. Consider a supply chain with $\sigma_\xi >> \sigma_\epsilon$. For this supply chain, the inefficiency due to the lack of credible forecast information sharing would be large because the supplier's knowledge of market demand is much less certain than that of the manufacturer's. One possible measure of *degree of forecast information asymmetry* is the ratio of the standard deviations $\frac{\sigma_\xi}{\sigma_\epsilon}$.

Özer and Wei (2006) show that the supplier and the manufacturer can choose among structured agreements that enable a mutually beneficial partnership depending on the risk adjusted profit margin and the degree of forecast information asymmetry. The results are summarized in Figure 9. For example, when forecast information between the parties is highly imbalanced, and the risk adjusted profit margin is high, then their analysis shows that the advanced purchase contract generates higher profits for both parties.

Through our private conversations with executives from several industries, we also observed that the (risk adjusted) profit margin and the degree of forecast information asymmetry are two primary

drivers of capacity risk. Figure 10 maps the level of these drivers for industries. For example, in the semiconductor industry, compared to the manufacturer, the supplier knows very little about the manufacturer's private forecast. Further empirical and field research is needed to verify Figure 10.
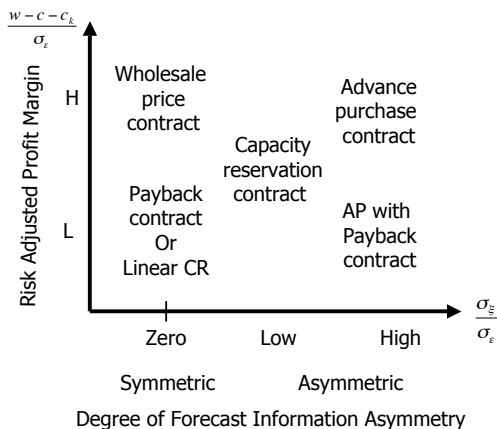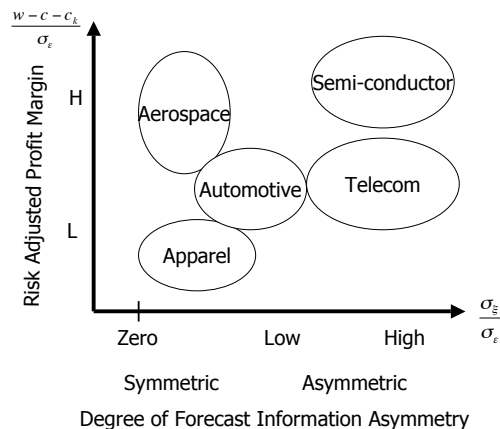

Figure 9: Mutually Beneficial Contracts

Figure 10: Capacity Risk Drivers across Different Industries

A Brief Review

Research exploring contracts that coordinate the supply channel under *full (or symmetric)* forecast information falls into two groups. In the first group, contracts align incentives by inducing the supplier and manufacturer to share the risk of low demand, resulting in excess capacity or inventory. Buyback contracts (Pasternack 1985), quantity flexibility contracts (Tsay 1999), and capacity reservation contracts (Erkoc and Wu 2001) are a few examples. The second category of contracts aligns incentives by sharing the risk of high demand, resulting in capacity or inventory shortage. Revenue sharing contracts (Cachon and Lariviere 2000) and quantity premium contracts (Tomlin 2003) are two examples from this category. Cachon (2003) provides a comprehensive review of supply chain contracting and coordination[9]. The supply chain literature that explicitly models *asymmetric information* can be classified into two groups. A group of researchers (Corbett and de Groote and Ha 2001) focus on information asymmetry in production cost, and another group (Porteus and Whang 1991, Cachon and Lariviere 2001, Özer and Wei 2006) focuses on information asymmetry in market demand and forecasts. Chen (2003) provides an excellent review of the use of these models in supply chains.

3.2 Inventory Risk

Lutze and Özer (2004) study the incentive problems in a multi-period, two-echelon supply chain with a manufacturer and a retailer both of whom build or procure to stock. Note that the manufacturer in this case faces inventory risk, unlike the previous section's build-to-order manufacturer. Both the manufacturer and the retailer hold inventory to satisfy their respective customers. They review inventory periodically, i.e., at the beginning of each period t. The manufacturer produces at a per

---

[9]In this literature, mainly the downstream firm is assumed to face demand uncertainty while the upstream firm "builds to order", unlike the interaction discussed in this section. Nevertheless, the results are analogous.

unit cost $c_m > 0$ and the retailer places an order at a per unit ordering cost $c_r > 0$. Suppose all cost and demand parameters are stationary, i.e., independent of period t. There is no fixed cost for production or placing an order. The manufacturer has ample capacity for production, which takes L periods to complete. The retailer orders are processed and shipped in l periods. Customer demand D_t is realized. The retailer satisfies customer demand through on-hand inventory. Unsatisfied demand is backlogged. Backorders of end customer demand incur a unit penalty cost p_r per period only at the retailer. The manufacturer incurs a shortage cost for unsatisfied retailer order based on the contractual agreement we specify later. The manufacturer and the retailer incur unit holding cost $h_m > 0$ and $h_r > 0$, respectively, where $h_m \leq h_r$, for any inventory remaining at the end of each period. Both the retailer and the manufacturer choose an optimal inventory replenishment policy to minimize their respective total expected inventory costs over T periods. At the end of period T, leftover inventory (resp., backlog) is salvaged (resp., purchased) at a linear per unit value of c_m and c_r, at each stage, respectively.

The manufacturer needs to protect himself against the retailer's demand over the production lead time L, and the retailer needs to protect herself against the consumer demand during the processing lead time l. Hence, to reduce inventory exposure, the manufacturer prefers the retailer to commit to purchase in advance and wait for delivery (commit and wait). However, the retailer prefers to delay her order and have immediate product availability and delivery (now or never). To address these opposing interests, Lutze and Özer consider a *promised lead time* contract with two parameters: promised lead time τ and corresponding per period lump-sum payment K.

Under a promised lead time contract, when the retailer places an order, the manufacturer promises to ship this order, in full, after τ periods. To guarantee this delivery, the manufacturer arranges an alternate sourcing strategy to fill retailer demand that exceeds the manufacturer's on-hand inventory. That is, the manufacturer *borrows* emergency units from an alternative source and incurs penalty p_m per unit per period until the alternative source is replenished.[10] The effect of promised lead time is to shift the responsibility for demand uncertainty from the manufacturer to the retailer. Note that if the retailer agrees to a promised lead time $L + 1$, exceeding the manufacturer's production lead time, the manufacturer builds to order for the retailer and does not carry any inventory.

Under this agreement, each firm independently solves a periodic-review inventory control problem discussed in § 2.1. Let x_t^j and y_t^j be firm $j \in \{m, r\}$ inventory position before and after ordering, respectively, in period t, where m and r stand for the manufacturer and the retailer. The following dynamic program recursion minimizes the cost of managing the inventory system for a finite horizon problem with $T - t$ periods remaining until termination.

$$J_t^j(x_t^j|\tau) = \min_{y_t^j \geq x_t^j}\{G^j(y_t^j|\tau) + \alpha E_D J_{t+1}^j(x_{t+1}^j|\tau)\}$$

where $J_{T+1}^j(x_{T+1}^j|\tau) \equiv 0$ for $j \in \{m, r\}$, and

$$G^m(y_t^m|\tau) = (1-\alpha)c_m y_t^m + E[h_m(y_t^m - D^{L+1-\tau})^+ + p_m(D^{L+1-\tau} - y_t^m)^+] \text{ and}$$
$$G^r(y_t^r|\tau) = (1-\alpha)c_r y_t^r + E[h_r(y_t^r - D^{l+1+\tau})^+ + p_r(D^{l+1+\tau} - y_t^r)^+].$$

---

[10]Similar alternative sourcing strategies are also discussed in Lee, So and Tang (2000) and Graves and Willems (2000).

For the stationary finite horizon inventory control problems, a myopic base stock policy is known to be optimal (Veinott 1965). These myopic base stock levels for the manufacturer and retailer are the minimizers of their respective single period cost functions and are defined as

$$Y^m(\tau) = F_{L+1-\tau}^{-1}\left(\frac{p_m - (1-\alpha)c_m}{h_m + p_m}\right) \text{ and}$$

$$Y^r(p_r, \tau) = F_{l+1+\tau}^{-1}\left(\frac{p_r - (1-\alpha)c_r}{h_r + p_r}\right).$$

Hence, with promised lead time τ, firm j orders up to an optimal base stock level $Y^j(\tau)$ if its inventory position x_t^j is below this level at the beginning of period t. The expected discounted inventory cost over T periods equals the sum of the discounted single period costs, that is,

$$J^m(x_1^m|Y^m(\tau), \tau) = \sum_{t=1}^{T} \alpha^{t-1}G^m(\tau),$$

where $G^m(\tau) \equiv c_m\mu + E\left\{h_m[Y^mt(\tau) - D^{L+1-\tau}]^+ + p_m[D^{L+1-\tau} - Y_t^m(\tau)]^+\right\}$, and

$$J^r(x_1^r|Y^r(p_r, \tau), \tau) = \sum_{t=1}^{T} \alpha^{t-1}G^r(p_r, \tau),$$

where $G^r(p_r, \tau) \equiv c_r\mu + E\left\{h_r[Y_t^r(p_r, \tau) - D^{l+1+\tau}]^+ + p_r[D^{l+1+\tau} - Y_t^r(p_r, \tau)]^+\right\}$.

When the manufacturer has full information about the retailer's inventory related costs, she can determine the optimal promised lead time contract (τ, K) by solving the following problem.

$$\begin{array}{ll}
\text{minimize}_{\tau,K} & \sum_{t=1}^{T} \alpha^{t-1}\{G^m(\tau) - K\} \\
\text{subject to} & K + G_r^*(p_r, \tau) \leq \pi_{max}^r \\
& \tau \in \{0, \ldots, L+1\}
\end{array} \tag{21}$$

The constraint ensures that the retailer is not charged a cost larger than her *maximum reservation cost*. Note also that the summation over T periods does not affect the solution of this problem, hence it can be dropped from the objective function for optimization purposes. The constraint must be binding at optimality. Otherwise, we can increase K and reduce the objective function. Substituting $K = \pi_{max}^r - G_r^*(p_r, \tau)$ one can solve for the optimal contract parameters.

To solve the above problem, the manufacturer needs to know the retailer's cost information. He can perhaps estimate h_r fairly accurately because he knows the value of the product. The same may not necessarily be true for p_r. Companies often state penalty cost as a strategic cost parameter never to be revealed. Lutze and Özer show that the retailer has every incentive to conceal her service level to end consumers (or equivalently the penalty cost structure). Intuitively, the retailer has an incentive to exaggerate the service level, thereby shortening the promised lead time for the same agreed upon price and reducing his expected inventory cost per period. Hence, it is often not possible for the manufacturer to know the retailer's penalty cost.

Suppose that there are two types of retailers in the market: one with a low penalty cost p_r^L and the other one with a high p_r^H. Suppose also that the belief is such that with probability q, she is a high penalty cost retailer and with probability $(1-q)$ she is a low penalty cost retailer.

To determine the optimal contract mechanism $\{(\tau^L, K^L), (\tau^H, K^H)\}$, the manufacturer solves the following problem.

$$\text{minimize}_{(\tau_i, K_i)_{i=L,H}} \quad q[G^m(\tau^H) - K^H] + (1-q)[G^m(\tau^L) - K^L]$$

$$\text{subject to}$$

$$IC_1: \quad K^H + G^r(p_r^H, \tau^H) \leq K^L + G^r(p^H, \tau^L)$$

$$IC_2: \quad K^L + G^r(p_r^L, \tau^L) \leq K^H + G^r(p_r^L, \tau^H)$$

$$IR_1: \quad K^H + G^r(p_r^H, \tau^H) \leq \pi_{max}^r$$

$$IR_2: \quad K^L + G^r(p_r^L, \tau^L) \leq \pi_{max}^r$$

$$\tau_i \in \{0, \ldots, L+1\} \text{ for } i = L, H$$

The first two incentive compatibility constraints ensure that a retailer with a high penalty cost voluntarily chooses the contract (τ^H, K^H) and the low penalty cost retailer chooses (τ^L, K^L). The next two individual rationality constraints guarantee the retailer finds a satisfactory contract regardless of his service level. This problem can be solved once we show certain properties of the cost function G^r and the result is in closed form solution. For example, IC_2 and IR_1 imply that IR_2 is redundant when we show G^r is increasing in p_r. Note also that IC_2 must be binding at optimality otherwise the manager can increase K^L and reduce the objective function until IC_2 binds. By showing that $G^r(p_r, \tau)$ has *single crossing property*[11], we can also show that $\tau^H \leq \tau^L$. Intuitively, it is optimal to offer a shorter promised lead time to a retailer that has higher penalty cost. Together with this observation, the binding IC_2 implies that IC_1 is redundant. At optimality

$$\tau^H = \text{minimizer of } q[G^m(\tau^H) + G^r(p_r^H, \tau^H)] + (1-q)[G^r(p_r^H, \tau^H) - G^r(p_r^L, \tau^H)]$$

$$K^H = \pi_{max}^r - G^r(p_r^H, \tau^H)$$

$$\tau^L = \text{minimizer of } (1-q)[G^m(\tau^L) + G^r(p_r^L, \tau^L)]$$

$$K^L = [\pi_{max}^r - G^r(p_r^L, \tau^L)] - [G^r(p_r^H, \tau^H) - G^r(p_r^L, \tau^H)]$$

Lutze and Özer (2004) discuss properties of optimal promised lead time contracts and the resulting inventory levels under both full and asymmetric service information with multiple discrete types. We caution that the results for mechanism design problems with multiple discrete types do not simply follow from two-type case. The study of the more general case requires intricate analysis and may lead to different solutions. Lovejoy (2006)'s paper is an excellent reference that clarifies related issues. Lutze and Özer (2004) also show how the ensuing inventory risk sharing strategy changes under asymmetric service information. They also compare the performance of a supply chain operating under a central decision maker to one with independent firms operating under a promised lead time contract. By comparing different control mechanisms and information scenarios, they provide insight into stock positioning and how the promised lead time affects the system performance. They quantify, for example, how much and when the manufacturer and the retailer over- or under-invest in inventory as compared to centralized supply chain, which operates as a serial system, discussed in § 2.5 and later in § 4.

---

[11]When $f(x, y) - f(x, y-1)$ is increasing in x, then function f is said to satisfy single crossing property.

3.3 Quality Risk

So far, we discussed the two risk imbalances in supply chains, leading to incentive problems. They were namely capacity and inventory risk. Next we discuss the third one: quality risk. The quality literature in operations management focuses mainly on centralized inventory management problems with random yields. Yano and Lee (1995) provide a review of more than 70 academic papers, such as the works of Porteus (1990b) and Pentico (1994). The focus of these papers is on establishing production and stocking policies when production or procurement yields are random. They address, for example, optimal time and size for inspection. Only a handful of researchers study the effect of product quality in decentralized supply chains. In Reyniers and Tapiero (1995), the supplier determines the effort invested in quality, where high effort causes a lower probability of defect. The manufacturer decides whether to conduct costly inspection. Lim (2001) uses a similar setting, with asymmetric information on the supplier's quality type. Baiman, Fischer and Rajan (2000) analyze the effects of different assumptions regarding the contractibility of quality and inspection efforts. All of these papers define quality as the percentage of the products that are not defective.

Today, manufacturers are outsourcing advance functions such as strategic sourcing, design, and even research and development. The manufacturer can use inspection techniques to measure yield and, hence, can enforce a certain yield in the contract. However, when the supplier undertakes more advanced tasks, measuring either the supplier's quality effort or his cost to achieve the desired quality level is difficult. This difficulty precludes the manufacturer from enforcing the desired quality level with a legal contract.

Not being able to foresee all possible contingencies and time to market pressures are two other reasons that make quality difficult to measure. Quality requirements may be better understood after the supplier builds a prototype, but this step typically occurs after an outsourcing agreement is signed. According to a Toshiba manager, if Toshiba waited until they were absolutely sure of every final detail and then wrote a complete contract, they would be 6 to 12 months late to the marketplace. Therefore, in addition to structured and legally binding agreements, establishing strategic relationship management systems between the manufacturer and the supplier is probably a good idea. This strategic relationship may encourage, for example, implementation of quality programs such as TQM or Six Sigma.

Kaya and Özer (2004) refer to the adverse effect of inefficiencies caused by the immeasurability of both quality effort level and the quality cost as the quality risk. Consider an original equipment manufacturer that outsources the design and production of a custom component to a supplier and sells the final product at a price p. The market demand is a function of the manufacturer's sales price p, the supplier's quality effort e and the market uncertainty ϵ, i.e.,

$$q = a - bp + e + \epsilon,$$

where $a > 0$ and $b > 0$ are the intercept and slope of the downward sloping demand curve. The manufacturer offers a procurement contract to the supplier. If the supplier accepts the contract, the parties establish a supply chain. Next, the supplier determines the product's quality level e by exerting costly quality effort. The quality cost is the supplier's private information. The manufacturer determines the sales price p to maximize her expected profit. The market shock ϵ realizes and the firms observe the quantity demanded. Finally, the supplier produces to satisfy the manufacturer's order, which is equal to the consumer demand. Note that the manufacturer

cannot *verify* the quality level set by the supplier due to the market uncertainty ϵ. Hence, the manufacturer cannot directly link the CM's compensation to the quality level the CM sets. Instead, the manufacturer needs to offer a contract and indirectly influence the CM's quality decision. The authors model this interaction as *screening* and *moral hazard* problems embedded into a three stage game.

Kaya and Özer (2004) design procurement contracts that improve the supplier's and the manufacturer's profits by inducing the supplier to exert effort to produce better quality products when parties cannot explicitly contract on quality. The authors answer and quantify broad questions of managerial interest. They quantify the value of being able to contract on quality. They study the effects of the manufacturer not knowing the supplier's cost of quality. They investigate the value of an enterprise-wide quality management system, a recent information technology tool that enables accounting of quality related activities across the supply chain. The authors also study the effect of the manufacturer's product-pricing policy on the resulting quality of the product. They report the outcome of two opposing product-pricing strategies: setting market price for the final product in the contract terms with suppliers versus pricing the product after receiving components from the supplier.

4 Large Scale Systems and Rationality

Global supply chains (or perhaps networks) have multiple locations to carry inventory; multiple products to manage; several decisions to coordinate; various sources and flows of information; and uncertain demand and processes. The management of inventory and information in such systems is difficult, and reviewing the related literature is even more so! We refer the reader to the books by Zipkin (2000) and Muckstadt (2005) for a systematic treatment of fundamental inventory control methods. First note that essentially multi-location systems or multi-product systems are identical. The same model and analysis can be applied to both. Second, by allocating decoupling inventories,[12] complex supply chain structures can be decomposed into fundamental structures, such as serial systems, of which we have a very good understanding. What is the best (if not the efficient) way to decompose a complex structure into smaller problems is an open research question. There is also an extant literature on the supply chain configuration problem (see, for example, Graves and Willems 2003). Here we will provide some discussion on how to allocate inventory effectively across two fundamental structures, serial and distributions systems, to minimize inventory related costs while keeping an eye on *rationality*.

Despite considerable progress over the years, existing optimization and policy evaluation algorithms for multi-echelon systems remains fragmented and opaque to non-experts. The computational methods involved are intricate and require voluminous data. Data fed to these tools are not always accurate, as discussed in § 2.3. Systems and people have limitations. Users are more likely to embrace decision tools when they understand what is in the black box. Therefore, it is necessary to develop *easy-to-describe*, *close-to-optimal* and *robust* heuristics that can be implemented on a spreadsheet by solving, for example, newsvendor type problems[13]. Unlike multi-echelon results, the

---

[12]Decoupling stock is used to permit separation of inventory decisions at different locations in the supply chain. Having a large inventory between two locations would make it possible for the downstream location to make an inventory decision independent of any supply problem at the upstream location.

[13]This problem is a simple single period, single location inventory control problem faced by a newsvendor. The

newsvendor problem is widely known, commonly used in practice and a standard component of any production and operations curriculum.

The above discussion suggest that heuristics and approximations can collectively enable better inventory management if they pass all or some of the following tests: (1) Is it close to optimal? (2) Is it simple to describe and use? (3) Can it be used to test system design issues accurately? (4) Is it robust? (5) Is it computationally easy? Note, however, that focusing narrowly on the one criterion overlooks other important aspect and leads to a gap between theory and practice (see Özer and Xiong 2005 for more discussions). For example, the computational methods used for exact solutions can be intricate and may require voluminous data. They may require advance knowledge. They may not provide explicit information regarding the key factors that drive performance. Recently, researchers have realized this gap and started to focus on developing easy-to-use, robust heuristics and approximations that are insightful (see, for example, Lee, Billington and Carter 1993, Hopp, Spearman and Zhang 1997, Gallego, Özer and Zipkin 2007, Shang and Song (2003), Gallego and Özer 2004, Caglar, Li and Simchi-Levi 2004, Watson and Zheng 2005, Özer and Xiong 2005 and references therein). In the following two subsections, we provide some examples from Gallego and Özer (2004) and Gallego, Özer and Zipkin (2007).

4.1 Serial Systems

Consider a serial system consisting of J stages. Stage $j < J$ procures from Stage $j + 1$ and Stage J replenishes from an outside supplier with ample stock. Customer demand occurs only at Stage 1 and follows a (compound) Poisson process, $\{D(t), t \geq 0\}$ with arrival rate λ. It takes L_j units of time for a unit to arrive at Stage j once it is released by its predecessor. Unsatisfied demand is backordered at each stage, but only Stage 1 incurs a linear backorder penalty cost p, per unit, per unit of time. We assume, without loss of generality, that each stage adds value as the item moves through the supply chain, so echelon holding costs h_j^e are positive. The local holding cost for stage j is $h_j \equiv \sum_{i=j}^{J} h_i^e$. The system is operated under continuous review. The following random variables describe the state of Stage j in equilibrium: D_j is the leadtime demand, I_j the on-hand inventory, and B_j the backorders. The total long-run average cost for any policy can be expressed as

$$E\left[\sum_{k=1}^{J} h_k I_k + p B_1 + \sum_{k=2}^{J} h_k D_{k-1}\right].$$

Optimality of an echelon base-stock policy $(s_J, \ldots, s_1)$ for this serial system is well known (see the original work by Clark and Scarf 1960). Gallego and Özer (2004) provide the following *new* recursive algorithm to obtain optimal base stock levels. Let $c_1(s) = E[h_1(s - D_1)^+ + p(D_1 - s)^+]$ and for $j = 2, \ldots, J$ define

$$
\begin{aligned}
c_j(s) &= \min_{x \in \{0, \ldots, s\}} c_j(x; s) \\
c_j(x; s) &= E[h_j(x - D_j)^+ + c_{j-1}(\min(s - x, s - D_j)) + h_j D_{j-1}].
\end{aligned}
\tag{22}
$$

Let $\mathcal{N} = \{0, 1, \ldots, \}$ be the set of non-negative integers and let

$$s_j^* \equiv \min\{s \in \mathcal{N} : c_j(s+1) - c_j(s) > h_{j+1}\} \text{ for } j = 1, \ldots, J.$$

vendor has to decide how much to order from the publisher so as to satisfy uncertain demand. The model is used to teach the risk of over-stocking and under-stocking

Function $c_j(s)$ is the long-run average cost of optimally managing the sub-system $\{j, \ldots, 1\}$ given echelon base-stock level s and s_j^* is the optimal base-stock levels. The recursion is somewhat intuitive. Suppose $c_j(\cdot)$ has been computed and consider the sub-system $\{j+1, \ldots, 1\}$. The goal is to compute $c_{j+1}(\cdot)$ from the knowledge of $c_j(\cdot)$. Note the link between the two sub-systems. We allocate x units to Stage $j+1$ and the remaining $s-x$ units of echelon base-stock to sub-system $\{j, \ldots, 1\}$. Given this allocation, the net inventory at Stage $j+1$ will be $(x-D_{j+1})^+$ which accrues at cost rate h_{j+1}. Since Stage $j+1$ will face a shortage when $D_{j+1} - x > 0$, the effective echelon inventory for sub-system $\{j, \ldots, 1\}$ is $s-x-(D_{j+1}-x)^+ = \min(s-x, s-D_{j+1})$. Thus, a finite local base-stock level at Stage $j+1$ imposes an externality to the sub-system $\{j, \ldots, 1\}$ whose expected cost is now $Ec_j(\min(s-x, s-D_{j+1}))$. As a result, when we allocate $x \le s$ units of local base-stock level to Stage $j+1$, the cost of managing a serial system with $j+1$ stages is given by (22).

Note that the classical recursive formulation presented in Chen and Zheng (1994) or Gallego and Zipkin (1999) has no intuitive interpretation. Although the above new algorithm has an interpretation and is intuitive, it is still difficult to explain to non-experts. It also does not provide any transparent relationship. Using this formulation, Gallego and Özer provide a fast exact algorithm based on gradient updates and a close-to-optimal heuristic that requires solving one newsvendor problem per stage! The heuristic is based on the approximate holding cost rate

$$h_j^{GO} \equiv \sum_{k=1}^{j} \frac{L_k}{L_1 + \ldots + L_j} h_k.$$

The idea is based on adding the holding cost as the product goes through the stages without delay and then dividing by the total lead time that it spends before reaching the end customer. This approximate holding cost minus the cost associated to upstream operations that is h_{j+1}, is charged to any excess inventory in echelon j that faces demand uncertainty over the leadtime $L_0 + \ldots + L_j$. This cost is charged per excess inventory because it is the approximate value that the echelon j is responsible for. Similarly the penalty cost $p + h_{j+1}$ is charged to echelon j because it is the approximate opportunity cost. The resulting problem then has a newsvendor type cost structure of

$$\tilde{c}_j(s) = E[(h_j^{GO} - h_{j+1})(s - \sum_{k=1}^{j} D_k)^+ + (p + h_{j+1})(\sum_{k=1}^{j} D_k - s)^+], \tag{23}$$

$$s_j^{GO} \equiv \min\{s \in \mathcal{N} : Pr(\sum_{k=1}^{j} D_k \le s) > \frac{p + h_j^{GO}}{p + h_{j+1}}\}. \tag{24}$$

Gallego and Özer (2004) show that over 1000 experiments, the optimality gap[14] is less than 0.25%. Note that the newsvendor problem is known to be somewhat robust in that small changes in data would not change the optimal solution significantly. This heuristic can easily be implemented with a simple spreadsheet. So, the heuristic is close-to-optimal, easy-to-describe and robust.

The authors also consider an approximation by approximating the leadtime demand distribution using Normal with mean μ and σ. This approximation would be good in particular when the mean of leadtime demand is large. The resulting cost $c_j(s_j^*) \approx (p + h_j^{GO})\sigma\phi(z)$, where $z = \Phi^{-1}((p + h_{j+1})/(p + h_j^{GO}))$. The base stock level is $s_j^* \approx \mu + z\sigma$. They are both in closed form. The authors also provide a distribution free cost upper bound in the approximate sense, that is $c_J(s_J^*) \le$

---

[14]The gap is defined as percentage difference between the optimal cost and the cost of the heuristic.

$\sqrt{p(h_1 L_1 + \ldots + h_J L_J)\lambda}$. The bound does not depend on any distribution. So it is quite robust with respect to demand parameter estimation. Such bounds, heuristics and approximations can also be used to quantify the value of system design issues. Using these results, it is easy to show, for example, that management should focus on reducing the lead time at the upstream stages while reducing the holding cost at the downstream stages. If process re-sequencing is an option, the lowest value added processes with the longest processing times should be carried out sooner rather than later. More importantly, a manager can easily quantify the impact of such changes using these simple heuristics and bounds.

4.2 Distribution Systems

Consider a two-level distribution system. All items enter the system from an external supplier and proceed first to location $j = 0$, called the warehouse. The warehouse in turn supplies J retailers, where the customer demands occur, indexed by $j = 1, \ldots, J$. Shipments from the external supplier arrive at the warehouse after time L_0. Shipments arrive at retailer j after time L_j. The retailers satisfy the customer demand from on-hand inventory, if possible. Unsatisfied demand at retailer j is backordered at a linear penalty cost rate b_j. All locations are allowed to carry inventory. The local holding cost is h_j per unit at retailer j. Holding inventory at the retailer is more expensive than holding it at the warehouse $h_j \geq h_0$ for $j > 0$. On the other hand, inventory located closer to the customer enables a quick response, hence reduces the possibility of a backorder at each retailer. Demand at each retailer j follows a Poisson process $\{D_j(t), t > 0\}$ with rate λ_j, and these are independent across retailers. The problem is, where to locate the inventory and how to control the system, so as to minimize the long-run average holding and penalty costs. No one knows the optimal policy for distribution systems, yet.

Under any policy, the total average cost can be expressed as

$$h_0 E[I_0] + h_0 \sum_{j>0} E[IT_j] + \sum_{j>0} (h_j E[I_j] + b_j E[B_j]),$$

where I_j is the on hand inventory, B_j is backorders and IT_j is the inventory in transit at location j at equilibrium.

Gallego, Özer and Zipkin (2007) distinguish two modes of control: central and local. Under central control, all information flows to one point, where all decisions are made. Local control means that each location observes local information and makes decisions accordingly. However, even under local control, a single decision-maker provides operating rules to all locations, which the locations then implement in real time. The locations do not have their own distinct objectives, as they do in contracting models of § 3.

Here we focus on the local control case and on a class of simple replenishment policies, base-stock or one-for-one policies (see Gallego, Özer and Zipkin 2007 for central control policies). Under local policy, whenever the inventory position at location j falls below the local base-stock level s_j, the retailer orders from the upstream location to raise the inventory position up to s_j. The sum of the retailers' orders constitutes the warehouse's demand process. The warehouse satisfies the retailers' requests on a first-come-first-served basis. Notice that information and control are decentralized or localized, in that each location sees its own demand and monitors its own inventory-order position. The exact analysis of this system is due to Simon (1971) and Graves (1985). Graves (1985) derives

the steady state distributions of inventory levels and backorders by disaggregating the backorders at the warehouse. Axsäter (1990) provides a recursive method to calculate the average holding and penalty cost associated with every supply unit that is matched with the demand that triggers it. The following random variables describe the system at equilibrium.

$$
\begin{align}
B_0 &= [D_0 - s_0]^+, \tag{25} \\
I_0 &= [s_0 - D_0]^+, \tag{26} \\
B_j &= [B_{0j} + D_j - s_j]^+ \text{ for } j > 0, \tag{27} \\
I_j &= [s_j - B_{0j} - D_j]^+ \text{ for } j > 0. \tag{28}
\end{align}
$$

Here, B_{0j} and D_j are independent, and $(B_{0j}|B_0)$ is binomial with parameters B_0 and $\theta_j = \lambda_j/\lambda_0$. Given the s_j, one can compute the $E[I_j]$ and $E[B_j]$ and thus

$$
\begin{align}
c(s_0, s_1, \ldots, s_J) &= h_0 E[I_0] + \sum_{j>0} c_j(s_0, s_j), \tag{29} \\
c_j(s_0, s_j) &= h_j E[I_j] + b_j E[B_j]. \tag{30}
\end{align}
$$

Let $s^* = (s_j^*)_{j=0}^J$ denote the policy that achieves the minimum average cost c^*.

For fixed s_0, the total average cost in (29) separates into a constant, plus functions c_j of one variable each (s_j), each convex in its variable. This separation is quite useful computationally. On the other hand, the remaining problem is still not trivial. To compute $E[B_j]$ and $E[I_j]$ requires numerical convolution of B_{0j} and D_j. Also, the cost $c(s_0, s_1^*(s_0), \ldots, s_J^*(s_0))$ is not convex in s_0. Finding the optimal s_0, therefore, requires an exhaustive search.

Gallego, Özer and Zipkin (2007) provide various heuristics based on restriction and decomposition ideas. Note, for example, that restricting the warehouse not to carry inventory decomposes the system to J retailers facing longer replenishment lead times, i.e. $L_0 + L_j$. This heuristic is referred to as cross-docking. To obtain the base-stock level at each retailer, they solve newsvendor type problems as in (23). The other extreme is to assume that the warehouse always has ample stock. Doing so, decomposes the system into individual retailers with lead time L_j. The authors solve for the warehouse's base-stock level by assuming that the retailers base stock levels are fixed to zero. In this case the warehouse's problem is a newsvendor type. The solution provides the maximum possible stock needed at the warehouse. Hence, they refer to this heuristic as stock-pooling. Another heuristic allocates zero safety stock to the warehouse, hence named the zero-safety heuristic. The authors show that a combination of these heuristics yields asymptotically optimal results, i.e. the combined heuristics yields optimal results as the number of retailer increases. Through an extensive numerical study involving all plausible distribution system parameters, the authors show that the optimality gap for the restriction and decomposition based heuristic is less than 2%. The authors also provide several other heuristics, bounds and approximations both for central and local control systems.

5 Ending Thoughts and Future Directions

We started with the discussion of how to effectively use information in centralized inventory systems. Such inventory systems are managed by a single decision maker who possesses all relevant

information. As we discussed in § 2, this line of research will always be necessary even though global inventory systems are decentralized in practice. Advances in technology, cheaper computational and storage devices will continue to enable managers to obtain more information. Inventory managers would need to quantify the value of information and the technology even more so than before. These systems also serve as a benchmark for decentralized systems. As we discussed in § 4, they are the building blocks for large scale systems. How to use and quantify new information in inventory management will continue to be an important area for future research.

Note also that there are still open questions. For example, we don't know the impact of imperfect inventory information on multi-echelon inventory systems. Intuitively, the adverse affect of inventory record inaccuracy will amplify as we go up in the echelon. Perhaps RFID technology has more value in such systems. But we simply don't know. Another example is the centralized distribution system, for which we still don't have an optimal inventory policy. As discussed in § 2, researchers have realized that an optimal policy would be very complex, if one exists. Hence, they have developed close-to-optimal heuristics, but none of these heuristics have worst case performance bounds. Developing such bounds is an interesting research direction. We have started to see recent research in this direction (Levi et al. 2006a,b).

Decentralized inventory management systems consist of managers with asymmetric information and separate objectives. We discussed several inefficiencies due to decentralized operations. Designing contracts to align incentives and coordinate inventory decisions will continue to be an important research area given global supply chains. Inventory managers need to keep an eye on inefficiencies introduced due to decentralization. Most of the work in this area consists of single period interactions between two inventory managers. Future work is needed to consider the effect of repeated interactions and reputation. This line of work also assumes that inventory managers are responsible for single location systems. In reality, however, an inventory manager could be responsible for a serial system or a distribution system (as discussed in sections 2 and 4). Hence, it is also important to study the interaction between two such managers. For example, the manufacturer in Lutze and Özer (2004) might offer a shorter promised lead time when he is managing a multi product inventory system due to perhaps the risk pooling effect. Studying the impact of supply chain design strategies such as postponement on contract terms would contribute to our understanding of these systems and bring us one step closer to real systems. The approximations, bounds and closed form solutions developed for centralized systems discussed in § 4 may also help us study complex decentralized inventory systems that are controlled by several managers.

Many years of research also suggest that large-scale, centralized stochastic inventory systems are even more difficult to deal with and are not amenable to a simple optimal policy. As a research community we need to develop close-to-optimal, easy-to-describe, robust heuristics for solving large scale systems. To make a heuristic universally acceptable, we need to test its performance against a lower bound or an optimal solution. For large scale systems, however, we lack optimal solutions. Developing sensible lower bounds could be difficult as well. As an alternative, such heuristics can be tested on real systems. However, real systems differ from each other, making it difficult to compare plausible heuristics proposed by researchers. Perhaps one potential research area is to design test-problems that are universally acceptable to qualify as difficult, real, and large-scale.

References

[1] Abhyankar, H.S. and S. Graves. 2001. Creating an inventory hedge for Markov-modulated Poisson demand: Application and model. *Manufacturing & Service Operations Management* 3 (4), 306 - 320.

[2] Alexander, K., T. Gilliam, K. Gramling, C. Grubelic, H. Kleinberger, S. Leng, D. Moogimane, C. Sheedy. 2002. Applying auto-ID to reduce losses associated with shrink. IBM Business Consulting Services, MIT Auto-ID Center White Paper, November.

[3] Anupindi, R., T. Morton and D. Pentico. 1996. The non-stationary stochastic lead time inventory problem: Near-myopic bounds, heuristics, testing. *Management Science* 42 (1), 124-129.

[4] Arrow, K. J. 1985. The economics of agency in principals and agents, J. P. Pratt and R. J. Zeckhauser (Eds.) Harvard Business School Press, Cambridge, MA.

[5] Atali, A. and Ö. Özer. 2005. Multi-Item inventory systems with Markov-modulated demands and production quantity requirements. *Working Paper*, Stanford University.

[6] Atali, A., H. Lee and Ö. Özer. 2004. Inventory control under imperfect information: Bounds, heuristics and approximations with demand prioritization. *Working Paper* Stanford University, Stanford, California.

[7] Atali, A., H. Lee and Ö. Özer. 2006. If the inventory manager knew: Value of visibility and RFID under imperfect inventory information. *Working Paper*, Stanford University.

[8] Aviv, Y. 2001. The effect of collaborative forecasting on supply chain performance. *Management Science* 47 (10), 1326-1343.

[9] Aviv, Y. and A. Federgruen. 2001. Design for postponement: a comprehensive characterization of its benefits under unknown demand distributions. *Operations Research* 49 (4), 578-598.

[10] Axsäter, S. 1990. Simple solution procedures for a class of two-echelon inventory problems. *Operations Research* 38, 64-69.

[11] Axsäter, S. 1993. Continuous review policies for multi-level inventory systems with stochastic demand. Chapter 4 in *Handbooks in Operations Research and Management Science* Vol 4, North-Holland.

[12] Azoury, K. S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science* 31, 1150-1160.

[13] Azoury, K. S. and B.S. Miller. 1984. A comparison of the optimal levels of Bayesian and non-Bayesian inventory models. *Management Science* 30, 993-1003.

[14] Baiman, S., P. E. Fischer and M. V. Rajan. 2000. Information, contracting and quality costs. *Management Science* 46, 776-789.

[15] Beyer, D. and Sethi, S. P. 1997. Average cost optimality in inventory models with markovian demands. *Journal of Optimization Theory and Applications* 92, 497-526.

[16] Benjaafar, S., W. L. Cooper and S. Mardan. 2005. Production-inventory systems with imperfect advance demand information and due-date updates. *Working paper.*

[17] Boyacı, T. and Ö. Özer. 2004. Information acquisition via pricing and advance sales for Capacity Planning: When to stop and act? *Working paper*, Stanford University.

[18] Buzacott, J.A. and J.G. Shantikumar. 1994. Safety stock versus safety time in MRP controlled production systems. *Management Science* 40, 1678-1689.

[19] Cachon, G. 2003. Supply chain coordination with contracts," Ch. 6 in *Handbook of Operations Research and Management Science* Vol. 11, Elsevier, Amsterdam.

[20] Cachon, G. and M. Lariviere. 2001. Contracting to assure supply: How to share demand forecasts in a supply chain. *Management Science* 47, 629-646.

[21] Cachon, G. and M. Lariviere. 2000. Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Forthcoming Management Science.*

[22] Caglar, D., C. Li and D. Simchi-Levi. 2004. Two-echelon spare parts inventory system subject to a service constraint. *IIE Transactions* 36, 655-666.

[23] Chen, F. 2003. Information sharing and supply chain coordination. Chapter 7 in *Handbooks in Operations Research and Management Science* Vol 11, Eds. A.G. de Kok and S. Graves. Elsevier, Amsterdam.

[24] Chen, F. 2001. Market segmentation, advanced demand information, and supply chain performance. *Manufacturing & Service Operations Management* 3, 53-67.

[25] Chen, F. and J. Song. 2001. Optimal policies for multi-echelon inventory problems with Markov modulated demand. *Operations Research* 49, 226-234.

[26] Chen, F. and B. Yu. 2005. Quantifying the value of leadtime information in a single-location inventory system. *Manufacturing & Service Operations Management* 7,144-151.

[27] Chen, F. and Y.-S. Zheng. 1994. Lower bounds for multi-echelon stochastic inventory systems. *Management Science* 40, 1426-1443.

[28] Clark, A. and H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Science* 6, 475-490.

[29] Cohen, M., T. Ho, J. Ren and C. Terwiesch. 2003. Measuring imputed costs in the semiconductor equipment supply chain. *Management Science* 49, 1653-1670.

[30] Cole, J. 1997. Boeing, pushing for record production, finds part shortages, delivery delays. *Wall Street Journal,* June 26.

[31] Corbett, C. J., G. A. DeCroix and A. Y. Ha, 2001. Optimal shared savings contracts in supply chains: linear contracts and double moral hazard. *Working Paper.*

[32] Diks, E.B. and A.G. de Kök. 1998 Optimal control of a divergent multi-echelon inventory system. *EJOR*, 111, 75-97.

[33] Dvoretzky, A., J. Keifer and J. Wolfowitz. 1952. The inventory problem: II. Case of unknown distributions of demand. *Econometrica* 20, 450-466.

[34] ECR Europe. 2003. Shrinkage: A collaborative approach to reducing stock loss in the supply chain.

[35] Eppen, G. and L. Schrage. 1981. Centralized ordering policies in a multi-warehouse system with lead times and random demand. *TIMS Studies in the Management Sciences* 16, 51-67.

[36] Erdhart, R. 1984. (s, S) policies for a dynamic inventory model with stochastic lead times. *Operations Research* 32, 121-132.

[37] Erkip, N., W. H. Hausman, S. Nahmias. 1990. Optimal centralized ordering policies in multi-echelon inventory systems with correlated demands. *Management Science* 36, 381-392.

[38] Erkoc, M. and D. Wu. 2005. Managing high-tech capacity expansion via reservation contracts. *Production and Operations Management* 14 (2) 232-251.

[39] Federgruen, A. 1993. Centralized planning models for multi-echelon inventory systems under uncertainty. Chapter 3 in *Handbooks in Operations Research and Management Science* Vol 4, North-Holland.

[40] Federgruen, A. and P. Zipkin. 1984a. Approximations of dynamic multilocation production and inventory problems. *Management Science* 30, 69-84.

[41] Federgruen, A. and P. Zipkin. 1984b. Allocation policies and cost approximations for multi-location inventory systems. *Naval Research Logistics Quarterly* 31, 97-129.

[42] Federgruen, A. and P. Zipkin. 1984c. Computational issues in an infinite horizon, multiechelon inventory model. *Operations Research* 32, 818-836.

[43] Files, J. 2001. Economic downturn leaves Cisco with stacks of excess inventory. *Knight-Ridder/Tribune Business News*, April 27.

[44] Fudenberg, D. and J. Tirole. 1991. Game theory. MIT Press, Cambridge, MA.

[45] Gallego, G. and Ö. Özer. 2001. Integrating replenishment decisions with advance demand information. *Management Science* 47, 1344-1360.

[46] Gallego, G. and Ö. Özer. 2001. Optimal use of demand information in supply chain management. Ch. 5 in Supply Chain Structures: Eds. J. Song and D. Yao. 2001, 119-160.

[47] Gallego, G. and Ö. Özer. 2002. A new algorithm and a new heuristic for serial supply systems. *Operations Research Letters* 33 (4), 349-362.

[48] Gallego, G. and Ö. Özer. 2003. Optimal replenishment policies for multiechelon inventory problems under advance demand information. *Manufacturing & Service Operations Management*, 5 (2), 157-175.

[49] Gallego, G., Ö., Özer and P. Zipkin. 2007. Bounds, Heuristics and Approximations for Distribution Systems. *Operations Research* 55(3), 503-517.

[50] Gallego, G. and P. Zipkin 1999. Stock positioning and performance estimation for multi-stage production-transportation systems. *Manufacturing & Service Operations Management* 1, 77-88.

[51] Gaukler, G., Ö. Özer and W.H. Hausman. 2004. Order progress information: Improved dynamic emergency ordering policies. To appear in *Production and Operations Management.*

[52] Gayon, J. P., S. Benjaafar and F. de Véricourt. 2007. Using imperfect demand information in production-inventory systems with multiple demand classes. To appear in *Manufacturing & Service Operations Management.*

[53] Graves, S. 1985. A multi-echelon inventory model for a repairable item with one-for-one replenishment. Management Science 31, 1247-1256.

[54] Graves, S., H. C. Meal, S. Dasu and Y. Qiu. 1986. Two-stage production planning in a dynamic environment. in *Multi-Stage Production Planning and Control.* Eds. S. Axsater, C. Schneeweiss, E. Silver. Springer-Verlag, Berlin, Germany, 9-43.

[55] Graves, S., D. B. Kletter and W. B. Hetzel. 1998. A dynamic model for requirement planning with application to supply chain optimization. *Operations Research* 46, S35-49.

[56] Graves, G. and S. Willems. 2003. Supply chain design: Safety stock placement and supply chain configuration. Ch. 3 in *Handbooks in Operations Research and Management Science* Vol 11, Eds A.G. de Kok and S. Graves. Elsevier, Amsterdam.

[57] Güllü, R. 1996. On the value of information in dynamic production/ inventory problems under forecast evolution. *Naval Res. Logist.* 43, 289-303.

[58] Ha, A. 2001. Supplier-buyer contracting: Asymmetric cost information and cutoff level policy for buyer participation. *Naval Research Logistics* 48, 41-64.

[59] Hariharan, R. and P. Zipkin. 1995. Customer-order information, leadtimes and inventories. *Management Science* 41, 1599-1607.

[60] Hausman W. H. 1969. Sequential decision problems: A model to exploit existing forecasts. *Management Science* 16, B93-B111

[61] Heath, D. and P. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions* 26, 17-30.

[62] Hibbard, J. 2003. The case of the obsolete inventory. *Red Herring* 123, 34-38.

[63] Hopp, W., M. Spearman and R. Q. Zhang. 1997. Easily implementable inventory control policies. *Operations Research* 45, 327-340.

[64] Hopp, W., R. Q. Zhang and M. L. Spearman. 1999. Easily implementable hierarchical heuristic for a two-echelon spare parts distribution system. *IIE Transactions* 31, 977-988.

[65] Hu, X., I. Duenyas, R. Kapuscinski. 2004. Advance demand information and safety capacity as a hedge against demand and capacity uncertainty. *Working paper.* University of Michigan.

[66] Iglehart, D. and R. Morey. 1972. Inventory systems with imperfect asset information. *Management Science* 18, B-388-394.

[67] Janakiraman G. and R. O. Roundy 2004. Lost-sales problems with stochastic lead times: Convexity results for base-stock policies. *Operations Research* 52, pp. 795-803.

[68] Iida, T. 2001. The infinite horizon non-stationary stochastic multi-echelon inventory problem and near myopic policies.*Eur. J. of Operational Research* 134(3), 525-539.

[69] Johnson, O. and H. Thompson. 1975. Optimality of myopic inventory policies for certain depended processes. *Management Science* 21, 1303-1307.

[70] Kaminsky, P. and J. Swaminathan. 2001. Utilizing forecast band refinement for capacitated production planning. *Manufacturing & Service Operations Management* 3, 68-81.

[71] Karaesmen F., J.A. Buzacott and Y. Dallery. 2002. Integrating advance order information in make-to-stock production systems. *IIE Transactions*, 34 (8), 649-662.

[72] Kambil, A. and J. Brooks. 2002. Auto-ID across the value chain: From dramatic potential to greater efficiency and profit. www.autoidcenter.com

[73] Kang, Y. and S. Gershwin. 2005. Information inaccuracy in inventory systems-stock loss and stockout. *IIE Transactions* 37, 843-859.

[74] Kaplan R. 1970. A dynamic inventory model with stochastic lead times. *Management Science* 491-507.

[75] Kaya, M. and Ö. Özer. 2004. Quality risk in outsourcing: Noncontractible product quality and private quality cost information. *Working Paper*, Stanford University.

[76] Lee, H., C. Billington and B. Carter. 1993. Hewlett-Packard gains control of inventory and service through design for localization. *Interfaces* 23, 1-11.

[77] Lee, H. and Ö. Özer. 2007. Unlocking the value of RFID. *Production and Operations Management* 16(1).

[78] Lee, H., P. Padmanadbhan and S. Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Science* 43, 546-558.

[79] Lee, H. and C. Tang. 1997. Modeling the costs and benefits of delayed product differentiation. *Management Science* 43, 40-53.

[80] Levi, R., M. Pal, R. Roundy and D. Shmoys. 2006a. Approximation algorithms for stochastic inventory control models. To appear *Math of Oper. Res.*

[81] Levi, R., R. Roundy and V. A. Truong. 2006b. Provably near-optimal balancing policies for multi- echelon stochastic inventory control models. *Working paper*.

[82] Luss, H. 1982. Operations research and capacity expansion problems: A survey. *Operations Research* 5, 907-947.

[83] Lütze, H. and Ö. Özer. 2003. Multi-product manufacturing and market segmentation with advance demand information. *Working paper*, Stanford University.

[84] Lütze, H. and Ö. Özer. 2004. Promised lead time contracts under asymmetric information. To appear in *Operations Research*.

[85] Lim, W. S. 2001. Producer-supplier contracts with incomplete information. *Management Science* 47, 709-715.

[86] Lovejoy, W. 1990. Myopic policies for some inventory models with uncertain demand distributions. *Management Science* 6, 724-738.

[87] Lovejoy, W. 2006. Optimal mechanisms with finite agent types. *Management Science* 47, 264-279.

[88] Marklund, J. 2006. Controlling inventories in divergent supply chains with advance-order information. *Operations Research* 54, 998-1010.

[89] Miller, L. M. 1989. Scarf's state reduction method, flexibility, and a dependent demand inventory model. *Operations Research* 34, 83-90.

[90] Moinzadeh, K. and C. Schmidt. 1991. An (S-1,S) inventory system with emergency orders. *Operations Research* 39, 308-321.

[91] Moinzadeh, K. and P. Aggarwal. 1997. An information based multiechelon inventory system with emergency orders. *Operations Research* 45, 694-701.

[92] Morton, T. and D. Pentico. 1995. The finite horizon non-stationary stochastic inventory problem: Near-myopic bounds, heuristics, testing. *Management Science* 41 (2), 334-343.

[93] Muckstadt, J. 1973. A model for a multi-item, multiechelon, multi-indenture inventory system. *Management Science* 20, 472-481.

[94] Muckstadt, J. 2005. Analysis and algorithms for service parts supply chain. *Springer*, New York, NY.

[95] Muharremoglu, A. and J. Tsitsiklis. 2003. A single-unit decomposition approach to multi-echelon inventory systems. *Working Paper*. Columbia University.

[96] Özer, Ö. 2003. Replenishment strategies for distribution systems under advance demand information. *Management Science* 49, 255-272.

[97] Özer, Ö. 2004. Capacity and quality risk in decentralized supply chains. *The Supply Chain Connection*, Vol 10, Issue 4.

[98] Özer, Ö. and W. Wei. 2006. Strategic commitment for optimal capacity decision under asymmetric forecast information. *Management Science*, 52,8, 1239-1258.

[99] Özer, Ö. and W. Wei. 2004. Inventory control with limited capacity and advance demand information. *Operations Research* 52, 988-1000.

[100] Özer, Ö. and H. Xiong. 2004. Stock positioning for distribution systems with service constraints. *Working paper*, Stanford University.

[101] Pasternack, B. 1985. Optimal pricing and return policies for perishable commodities. *Marketing Science* 4, 166-176.

[102] Pentico, D.W. 1994. Multistage production systems with random yield and recourse. *Int. J. Prod. Res* 32, 2455-2462.

[103] Porteus, E. 1990a. Stochastic inventory theory. Chapter 12 in *Handbooks in Operations Research and Management Science* 2, 605-652.

[104] Porteus, E. 1990b. The impact of inspection delay on process and inspection lot sizing. *Management Science* 36, 999-1007.

[105] Porteus, E. and S. Whang. 1991. On manufacturing/marketing interface. *Management Science* 37, 1166-1181.

[106] Raman, A., N. DeHoratious and Z. Ton. 2001. Execution: The missing link in retail operations. *California Management Review* 43, 136-152.

[107] Reyniers, D. J. and C. S. Tapiero. 1995. Contract design and the control of quality in a conflictual environment. *European Journal of Operational Research* 82, 373-382.

[108] Rinehart, R. F. 1960. Effects and causes of discrepancies in supply operations. *Operations Research*, 8(4) 543-564.

[109] Rosling, K. 1989. Optimal inventory policies for assembly systems under random demands. *Operations Research* 37, 565-579.

[110] Salanie, B. 1997. *The economics of contracts: A primer* MIT Press, Cambridge, MA.

[111] Scarf, H. 1959. The optimality of (s, S) policies in the dynamic inventory problem. Chapter 13 in *Mathematical Methods in the Social Sciences.* Eds. K. Arrow, S. Karlin, and P. Suppes, *Stanford University Press*, CA.

[112] Scarf, H. 1960. Bayes solutions of the statistical inventory problem. *Naval Logistics Research Quarterly* 7, 591-596.

[113] Schwarz, L. B. 1989. A model for assessing the value of warehouse risk-pooling: Risk-pooling over outside-supplier leadtimes. *Management Science* 35, 828-842.

[114] Schwarz, L. B., N. C. Petruzzi and K. Wee. 1998. The value of advance order information and the implication for managing the supply chain: An information/control/buffer portfolio perspective. *Working paper.* Purdue University.

[115] Sherbrooke, C. 1968. METRIC: A multi-echelon technique for recoverable item control. *Operations Research* 16, 122-141.

[116] Shang, K and J. Song. 2003. Newsvendor bounds and heuristics for optimal policies in serial supply chains. *Management Science* 49, 618-638.

[117] Simon, R. M. 1971. Stationary properties of a two-echelon inventory model for low demand items. *Operations Research* 19, 761-773.

[118] Simchi-Levi, D., P. Kaminsky and E. Simchi-Levi. 2000. Designing and managing the supply chain: concepts, strategies, and case studies. *Irwin/McGraw-Hill series, Operations and decision sciences.*

[119] Song, J. and P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Operations Research* 41, 351-370.

[120] Song, J. and P. Zipkin. 1996. Inventory control with information about supply condition. *Management Science* 42, 1409-1419.

[121] Tan, T., R. Güllü and N. Erkip. 2006. Modeling imperfect advance demand information and analysis of optimal inventory policies. To appear in *Europ. Jour. of Oper. Research*

[122] Toktay L. B. and L. W. Wein. 2001. Analysis of a forecasting-production-inventory system with stationary demand. *Management Science* 47, 1268-1281.

[123] Tomlin, B. 2003. Capacity investment in supply chains: Sharing-the-gain rather than sharing-the-pain. To appear in *Manufacturing & Service Oper. Management.*

[124] Tsay, A. 1999. The quantity flexibility contract and supply-customer incentive. *Management Science* 45, 1339-1358.

[125] Van Houtum, G. J., K. Inderfurth and W. H. M. Zijm. 1996. Materials coordination in stochastic multi-echelon systems. *EJOR* 95, 1-23.

[126] Veinott, A. 1965. Optimal policy for a multi-product, dynamic, non-stationary inventory problem. *Management Science* 12, 206-222.

[127] Veinott, A. 1966. On the optimality of (s, S) inventory policies: New conditions and a new bound. *J. SIAM Applied Mathematics* 14, 1067-1083.

[128] Watson, N. and Y. Zhang. 2005. Decentralized serial supply chains subject to order delays and information distortion: Exploiting real time sales data. *Manufacturing & Service Operations Management* 7, 152-168.

[129] Yano, C. and H. Lee. 1995. Lot sizing with random yields: A review. *Operations Research* 43, 311-334.

[130] Zhu, K. and U. W. Thonemann 2004. Modeling the benefits of sharing future demand information. *Operations Research* 52, 136-147.

[131] Zipkin, P. H., Foundations of inventory management, McGraw Hill, Boston, 2000.