

INVENTORY CONTROL

Part One - Basic Inventory Concepts

The Context of Inventory Control

The Importance of Inventory

The design and operation of any logistics system or supply chain that deals with the manufacture and distribution of physical goods must address fundamental questions about how the inventory in the system will be managed and controlled. Over the past few decades, management attention has increasingly focused on the firm's inventory as an area where improvements can be made so as to both reduce costs and improve the level of customer service that the enterprise delivers to its customer base. Reduction in the inventory level maintained in the enterprise, and hence in total assets and investment, is seen as an important tool to improve the firm's rate of return on assets, return on investment, and ultimately, its shareholder value and stock price.

Aggregate levels of inventory have come to be seen by some as a kind of global metric of the firm's supply chain performance or efficiency. Corporations try to benchmark themselves against competitors and other firms, where low levels of inventory are seen as very desirable. While aggregate inventory is usually too simple and crude a measure of true supply chain performance, it is a highly visible component of the firm's asset structure, and because the level of inventory can be so easily manipulated (at least in theory), efforts to control and reduce inventory receive high-level attention.

Over the past thirty years, scores of new inventory control procedures have been introduced across many different industries and lines of trade. The trade press and academic literature contains thousands of accounts of tools such as EOQ, MRP, DRP, JIT, VMI, and the list goes on, seemingly forever. Each new procedure was intended to address a significant weakness with the current state of practice, and in many cases firms reported great improvement in their inventory operations when the new procedure was implemented. In almost every case, the new approach was intended to reduce the amount of inventory in the system while providing a level of support as least as good as, if not in fact superior to, that provided by the prior practice. While firms have been striving for decades to reduce the level of inventory in their systems, empirical evidence (cites) has shown that the actual level of aggregate inventory in the national economy, sector by sector, has remained stubbornly high. Inventory management, it seems, is not so simple after all. A decision maker will benefit from a comprehensive understanding of the issues and problems associated with inventory control in the supply chain, as well as from a fundamental understanding of the range of inventory control tools which are available, how

they work, and what problems they are designed to solve. The purpose of this section is to contribute to this understanding.

The Purposes of Inventory

Inventory defined. Inventory is material that the firm obtains in advance of need, holds until it is needed, and then uses, consumes, incorporates into a product, sells, or otherwise disposes of. A business inventory is by its nature temporary; that is, the firm expects to hold an individual item as inventory for a relatively brief period of time. However, due to the ongoing nature of the firm's operation, aggregate or total inventory is continually being reordered and restocked. The net effect of these ongoing processes of usage and replenishment is that the firm retains a "permanent investment" in inventory. In other words, individual units come and go, but the firm generally maintains some amount or level of total or aggregate inventory. These inventory assets can be as important to the financial structure of the firm as is investment in, for example, plant and equipment.

Inventory management and control consists of developing good answers to the following questions:

- A. Which items will the firm carry as inventory?
- B. Where in the logistics system will these inventories be maintained?
- C. How large will the inventory for each item be?

Inventory by its nature is dynamic over time. For a typical item with a long product lifetime, inventory stocks rise with replenishment actions and fall with usage or consumption. Decisions about the size of the inventory are therefore implemented over time by monitoring and adjusting the sizing and the timing of inventory replenishment quantities to bring about the desired average or aggregate inventory size. The larger issue here is to determine what an appropriate or correct inventory posture might be for the firm. To answer this question, the analysis must carefully consider:

- A. The benefit provided by the inventory. That is, what purpose does the inventory serve, and what is the value of that service?
- B. The total cost of the inventory. Many of these costs are obvious, but some others are not. An understanding of the true costs associated with inventory is essential.
- C. Other logistics options that could be used in place of inventory. For example, it may be possible to replace an investment in inventory with premium or high speed freight service.

A key insight to consider here is that inventory is can often be viewed as a cost-reduction tool. The real purpose of inventory is to reduce the total costs involved in the operation of the logistics system. There is usually a way to reduce the level of inventory in a system, but sometimes these alternatives are simply too costly. This is another way of saying that sometimes holding inventory is the cheapest and best way to solve a logistics problem.

Process Time Inventory. Process Time Inventory is needed because it takes time to perform some useful function on or with the item in question. For example, it takes some amount of time to ship goods from point to point in a distribution system. While these goods are being transported from one location to the next, they constitute in-transit or pipeline inventories. Raw materials and component parts often spend a considerable amount of time within a manufacturing facility while they are being transformed into products or finished goods. These Work In Process (WIP) inventories arise simply because the manufacturing process takes some minimum amount of time. In practice, these WIP inventories are often much large than would be implied by the manufacturing cycle times alone because components and assemblies “wait” at various stages in the manufacturing process due to scheduling considerations for multi-purpose manufacturing equipment or bottlenecks at highly utilized points of the facility. Process Time Inventory is necessary in most production and distribution systems, but it should be managed carefully to insure that the value provided by the inventory is worth the expense incurred.

Process Uncoupling Inventory. Another important purpose that inventory serves is to uncouple processes in the logistics system. There are two fundamental reasons to uncouple these processes: the first is to obtain economies of scale in the operation, and the second is to deal with the inherent randomness or uncertainty in the processes.

As an example of scale economies, consider the manufacturing process and the sales process in a firm. We might choose to closely couple these two activities by building a plant with a production capacity that produces exactly one day’s worth of sales every day. If the daily sales volume for this item is low, the required plant might be quite small and inefficient, and therefore high in total unit cost. Suppose instead that we build a much larger facility that can mass-produce a thirty-day supply of this item in one batch in a single day. We can use this one high-speed plant to produce a product range of thirty such items and keep up with their total demand because we have uncoupled the production rate for each item from its sales rate. Because of scale economies, the single large facility will presumably be more cost-effective than the thirty small facilities it would replace. The important point to note here is that it is our ability and willingness to hold up to a month’s worth of finished goods inventory for each item which allows us to capture the benefit of the scale economies presented by the single high-speed

facility. Inventory is the tool that allows us to uncouple the process rates and claim the resulting cost improvement.

In some cases the nature of the product is such that inventory cannot be used to uncouple production and usage rates. Consider the case of electrical power generation. Power plants can generate electricity most efficiently, and therefore at the lowest total cost, at one specific and level rate of output. Unfortunately, in most regions the demand for electric power fluctuates considerably based on time of year and time of day. The desirable approach would be to produce electricity at a constant rate, and "bank" or store the power during periods of low usage to handle peak usage requirements. The obvious problem here is that the only available storage technologies available (batteries, etc.) are prohibitively expensive. As a result, power companies are forced to use a set of other strategies, including an elaborate national power distribution grid, expensive stand-by generation equipment, as well as peak and off-peak pricing schemes, in an endless attempt to balance supply and demand in real time. In spite of these efforts, we still face brown-outs and rolling blackouts during periods of extremely high usage. This power management process would be vastly more easy to accomplish if an economical large-scale power storage technology were available, that is, if we could economically maintain a "finished goods inventory" of electrical power.

The second common reason to uncouple processes in a logistics system is to isolate and deal with inherent uncertainty in the processes. Typically, a production process is most efficient if it is allowed to operate at a smooth, steady rate. Customer demand for the good being produced is often highly random and difficult to forecast with great precision. The firm can therefore hold a finished goods inventory to isolate these two processes. Production can proceed at a constant, economically efficient rate independent of the current level of demand, while the highly variable demand process can be satisfied with high levels of customer service, because the inventory serves as a buffer between the processes.

The inherent uncertainty in a distribution system is not always on the demand side. For example, in a municipal water system the actual daily usage is often fairly stable and quite predictable. In many cases the water supply relies on replenishment based on rainfall, which is relatively sporadic and unpredictable. The natural response to the problem is to build reservoirs that allow us to uncouple a random supply process from a fairly steady demand process by holding an inventory.

Anticipation and Speculative Inventory. An extreme form of process uncoupling occurs in the use of "anticipation inventory" on strongly seasonal items. Many consumer goods, such as children's toys, gas barbecue grills, and seasonal greeting cards, are sold in very high volumes during a very brief selling period each year. Rather than build a factory that can produce at these high

demand rates, and which would be idle for the rest of the year, most manufacturers build a much smaller facility. They operate the plant at a fairly steady rate during most of the year to build the inventory they will need to support their short selling season. For this anticipation strategy to work, the item in question must be "storable" at a reasonable cost, the firm must be able to forecast the level of demand reliably, and the design of the product must be "frozen" up to a year in advance. Where these conditions are met, this strategy can dramatically reduce the costs of production.

Sometimes inventory is held on a speculative basis. If the price of a raw material is expected to rise in the near future, it may be sensible to place an unusually large order now to obtain a large supply at the current favorable price. Such a decision should be based on the cost of holding the extra inventory and the extent, duration, and likelihood of the price rise. As the name implies, speculative inventories represent a gamble or a bet the firm makes about the movement in prices. When the goods in question are true open market commodities, like agricultural and mineral products, futures contracts and hedging strategies are used to protect the firm from price volatility. While important, these financial strategies are beyond the scope of this text.

The Importance of Demand Forecasting

Since inventory is material obtained in advance of need, any inventory control policy must be based on some anticipation or belief about which materials will be needed in the future, how much will be needed, and when the need will arise. In other words, inventory control must involve some forecast of future demand, whether crude or sophisticated, explicit or implicit. Even a Kanban system involves some forecasting, in that the size or desired quantity in the Kanban container is based on a consideration of the amount of demand which could occur over the re-supply time interval.

In a very real sense, the ability to forecast demand and re-supply times accurately will set an upper limit on how successful or efficient the inventory control policy will be. In many cases, the most effective way to improve the inventory performance in a firm may not involve changing the inventory control algorithm, but rather would focus on improving the demand forecasts that drive the existing inventory system.

That having been said, there is far more to effective inventory control than simply forecasting demand "accurately", and it is often impossible to obtain truly accurate demand forecasts. An appropriate inventory control will accept the available inventory forecasts and operate as efficiently as is possible given the inherent inaccuracies of the forecasts.

Classifications of Inventory

Many different classification schemes are used to describe inventories and inventory systems within the firm. Some of the most widely used include:

Raw Materials / Work In Process / Finished Goods. Items are often categorized as either raw materials (or component parts), work in process, or finished goods as they enter, pass through, and leave a production or manufacturing process. For example, a furniture manufacturer may purchase three inch hex head bolts in bulk quantities from a supplier and hold them as a "raw materials" or component parts inventory until they are needed in an assembly process. At that point, some of the bolts are removed from the raw materials inventory and become incorporated into the "work in process" or "semi-finished good" inventory. When the assembly process is complete, the bolts enter the "finished goods" inventory, although the ability to track or trace the identity of the components is usually lost once the components are incorporated into the next higher assembly.

Inventory must be managed at all three stages in the manufacturing process to insure that sufficient supplies exist to support the customer demand, as well as the manufacturing schedule, without building unneeded and wasteful levels of inventory in any stage. In practice, different systems and techniques are often used in a firm, because the tools that are appropriate to manage a raw materials inventory will not necessarily be useful in managing the firm's finished goods. This functional definition of an item as raw materials or finished goods is always relative to a specific manufacturer in a supply chain. To the furniture manufacturer, the bolts are raw materials, but to the hardware supplier, the bolts are finished goods.

Cycle Stock and Safety Stock. When an item undergoes repetitive usage or demand over a long period of time, the firm usually establishes a periodic replenishment policy to maintain an inventory on the item. For example, the firm might choose to reorder a replenishment amount that is expected to be equal to sixty days worth of demand or usage of the item about every sixty days. This policy results in the establishment of a *cycle stock* that will vary in size from day to day as usage reduces the quantity on hand and replenishment increases it. Note that the choice of the order size, or the replenishment quantity, will determine the replenishment cycle, or *cycle time*, which is the average time between the receipt of replenishment orders, and will also determine the average number of inventory replenishment orders placed per year. With an order quantity equal to a "sixty day supply" of the item, the average size of the cycle stock will be about one half the order quantity, or thirty days worth. Thus the cycle stock will "turn" twelve times per year, since the annual demand will be twelve times the average held inventory. In other words, the size of the replenishment order will determine the *inventory turn rate* associated with the cycle stock. Later in this chapter we will explore in great detail a set of models

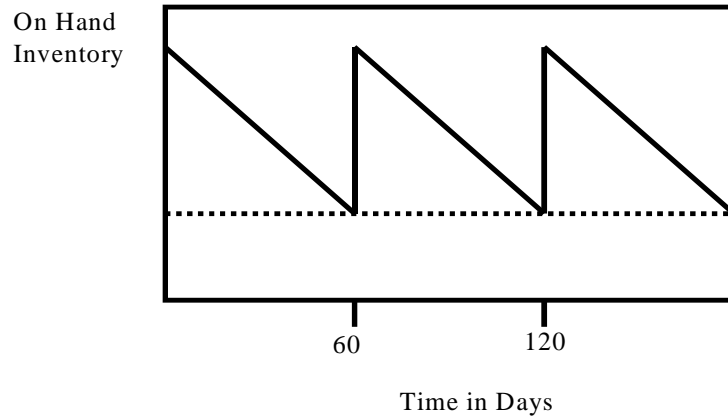
that will help to determine what an appropriate amount of cycle stock would be in a given situation.

In addition to the cycle stock, the firm may also maintain an additional amount of inventory, called *safety stock*, to deal with the potential uncertainty in the replenishment situation. For example, as a given inventory cycle comes to an end, a replenishment order is launched, and the *reorder lead-time* begins. If the resupply is later than expected, or if the actual demand during the replenishment lead-time is higher than had been anticipated, then the on-hand inventory would be completely depleted, and the inventory system would experience a *stockout* situation. Demands that occur during the stock out would typically become either *backorders* (in the case where the customer or user will wait until the resupply arrives) or *lost sales* (in the case where the demand is lost to the system because the user will not permit the backorder). In commercial and industrial practice, backorders are commonly permitted. At the retail customer level in a system, stockouts usually result in lost sales. In either case, the stockout situation results in poor customer service and the attendant costs. In order to reduce the number of stockouts, and the associated backorders or lost sales, the firm can carry safety stock.

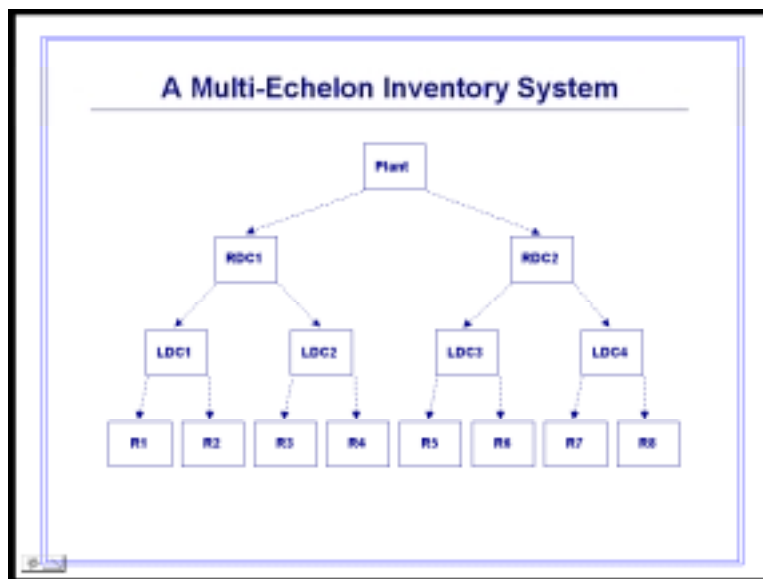
Note also that this safety stock represents a relatively “permanent investment” in inventory. It is only really needed at the end of a replenishment cycle, since that is where the stockout risk occurs. Unfortunately, to have it available at the end of the cycle, it must be available all through the cycle as well. What this means is that the average amount of safety stock on hand over time is very nearly constant. In the diagram below, we see an idealized representation (a so-called *inventory saw-tooth*) which shows how the on hand inventory level would rise and fall in a replenishment scenario with a highly consistent demand process. In this example, the chosen safety stock level is about half the order quantity; that is, it represents about a thirty-day supply. Thus the total inventory (cycle stock plus safety stock) would represent about sixty days worth of inventory on average, and the total inventory turn rate would therefore be about six times per year.

The safety stock acts as an insurance policy against the chance of stockouts. As such, the appropriate quantity of safety stock will depend on the relative risks and costs in a given situation. This chapter will also develop a set of probabilistic models that will help to determine a cost-effective safety stock strategy in a given situation.

Cycle Stock and Safety Stock



Multi-echelon Inventory Systems. Many firms will maintain an item in inventory at a large number of different locations throughout the system. For example, a large retailer will carry an identical item (or *stock keeping unit*, or SKU) in inventory at thousands of different retail locations spread across a continent or the globe. An inventory control system of this type is large, but need not be inherently more complex than a control system that manages a single item at a single location. Traditional cycle stock and safety stock models usually develop a solution, or a control strategy, for a single item at a single location. The procedure is simply repeated for every item at every location, generating rules for each item at each point.



But in a large distribution system there is usually another complicating factor. In order to capture economies of scale in the production process, the product is manufactured at one or a small number of plants, and it moves out to the thousands of retail locations by passing through a series of intermediate distribution inventories. Thus the inventory system consists, not simply of multiple locations, but rather of a series of inter-related layers, or tiers, or *inventory echelons* through which the stock moves on its way to the retail level.

In the diagram, for example, one production plant serves two regional distribution centers (RDC1 and RDC2), which serve in turn, let us say, the eastern and western regions of North America. The regional facilities order and receive inventory to satisfy orders from a series of local distribution centers (LDC1, LDC2, etc.) each of which is assigned to serve a portion of the region. The local facilities send inventory along to the retailers (R1, R2, etc.) in their zone in response to replenishment orders from the retailers. In this kind of system, the inventory follows a cascading or stair-step pattern on its journey through the system. For example, a specific unit of stock might be produced in January, held as a part of the finished goods inventory at the plant through February, be held as cycle stock at RDC1 in March, pass through LDC2 in April, and finally sit on the retail shelf at R4 in May.

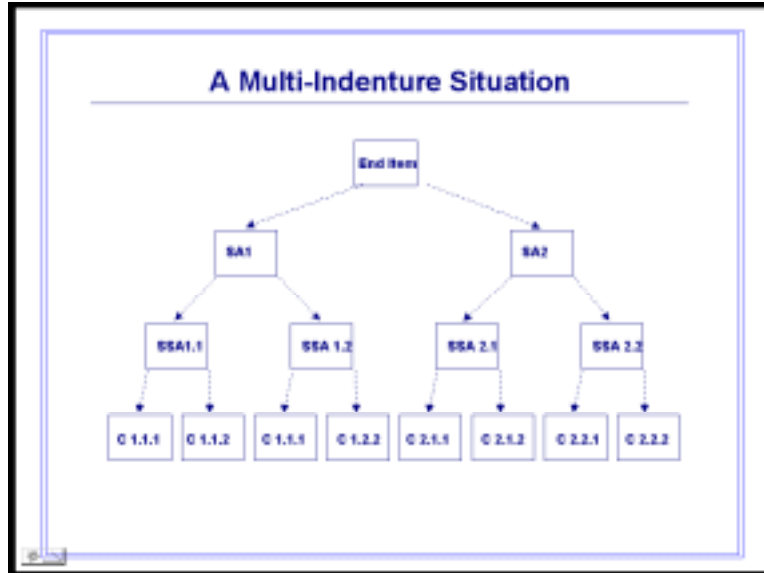
This may seem like an unnecessarily complicated and costly way to handle the distribution, since it involves extra facility costs, multiple material handlings, re-shipments, and multiple inventory positions for the item. Why don't we eliminate the intermediate echelons and simply ship directly from the plant to the retailers? In fact, some companies do just that, but a system of intermediate distribution inventories is more common for sound economic reasons. Any given retailer will place a relatively small sized order at any given time. The plant would be overwhelmed with thousands of very small orders, which would be very expensive to ship over the long distances to the retailers. By operating a system of well positioned intermediate stockage points, the firm can schedule and ship very economical full truckload, or full railcar load, or full ship load quantities at dramatically reduced rates. In addition, by pre-positioning the inventory in the local area, the firm can respond very quickly to a retail order when it does happen without needing to ship long distances using prohibitively expensive premium freight services. In other words, the scale economies in the transport of freight, combined with the market requirement to provide fast and reliable delivery times, drive most large firms to operate multi-echelon distribution inventories. In a well designed multi-level system, the layers of warehousing and inventory more than pay for themselves, and the system reduces, rather than increases, total cost.

In practice, however, multi-echelon inventory systems are much more difficult to direct and control than a "flat" or single-echelon system. Experience has shown that simple control procedures which are suitable for single level

systems break down badly in the multi-level environment. Over the last twenty years a number of inventory systems, most notably Distribution Resource Planning, or DRP, have been developed specifically to operate in this framework. While good progress has been made, multi-echelon systems remain a challenge both to the theory and practice of inventory control.

Multi-indenture Inventory Systems. Another common form of complexity which occurs in large-scale production-distribution systems arises when the many individual items in the system are inter-related through their demand or usage patterns. In most retail situations, demand across SKUs is reasonably uncorrelated. That is, the quantity that is sold or demanded of a given item on a given day would not be expected to correspond closely to the demand for another item. We would expect the demand processes for the items to be independent of one another, such that knowing the actual demand for SKU A today would not improve our ability to forecast today's demand for SKU B. In a situation like this, it can make sense to manage the retail inventory as a whole by simply treating each item as an independent inventory control problem to be solved. Most traditional cycle stock / safety stock systems operate in just this way.

However, a very different situation obtains in a typical production operation where a wide range of raw materials and component parts are held as inventory in support of the manufacturing process. Consider a situation where a



high speed, high volume manufacturing facility is used to produce a wide array of product end items. An end item is produced in large volumes in a batch run, and then the assembly line or production process is changed over to make the next item. While the end item is in production, a large quantity of component parts are needed to support production. When the production batch is completed, the

usage or demand rate against these parts may go to zero. The second end item may be composed of a completely different set of component parts. In this situation, demand for the potentially thousands of component SKUs is highly correlated – in fact, demand for the components is virtually completely predictable because it is completely determined by the planned production schedule and by the recipe of components, or the so-called *Bill of Materials*, which is the identification of the assemblies, sub-assemblies, components and sub-components which are used to build the end item. Each level of assembly to sub-assembly relationships in the Bill of Materials is referred to as an *indenture*, and in a technologically complex product the indenture structure can be thirty or more levels deep. The component and sub-assembly fabrication and assembly points will not necessarily be co-located in the same facility; they are frequently spread around the globe and positioned in the operations of supply chain partners, contract manufacturers, and industrial customers.

In this production environment, inventory tools which focus on the derived or dependent nature of component demand are needed to capitalize on the information provided by the firm's production planning process. The first large-scale system which was developed to do this was Material Requirements Planning, or MRP, which has been in a state of continuous development and implementation since the early 1970s. Early MRP systems, while they were quite simple and crude, often produced a great improvement in both efficiency and effectiveness compared to the more traditional cycle stock – safety stock systems that they superceded. A great deal of work has been done since to improve on the basic MRP approach, but the fundamental insight provided by exploiting the nature of derived demand in multi-indenture systems remains the bedrock approach to managing pre-production inventories.

Maintenance, Repair, and Operations (MRO) Stocks. Another source of inventory is stockpiles of materials which are not directly incorporated into a product but rather are needed to operate and maintain the production facility. Many of these items fall into the category of “mops and brooms” or “pencils and paperclips”. As such, they do not seem to provide much of a management challenge, and inventory control of these items is usually straightforward. On the other hand, in some cases there is surprising room for improvement in the management of these items. A great deal of supply chain management attention has been focused in this area over the last few years, but the emphasis has been on improved techniques for purchasing or sourcing these items, for example, through the use of internet-based auctions and exchanges.

There is an important category of MRO stocks, however, which does merit special attention, and that is the category of spare parts. In many cases the firm will operate a large set of installed equipment, such as motor vehicles, aircraft, network servers, and so forth. This set of hardware, the *installed base*, is often expensive and technologically complex. It is operated at geographically dispersed locations, and it must be operated with high reliability and system

availability. To maintain the installed base, an investment is made in a set a spare parts which are used to restore the equipment to serviceable condition after it has failed. These spare parts are often very expensive in themselves, and depending on the technology involved, can often have very short useful lifetimes. In some cases, the spare parts are also *recoverable*, that is, a failed unit can be sent to an overhaul facility, refurbished, and then returned to the spare parts inventory. Effective management of a large-scale recoverable spares operation can be further complicated because of the multi-echelon and multi-indenture nature of the spares process. A special set of inventory control techniques which are based some elements of queuing theory have been developed over the last three decades which directly address the unique characteristics of the recoverable spare parts inventory problem.

The Nature of Inventory Costs

In developing an inventory control procedure or algorithm, we will generally seek to either minimize the sum of relevant costs or maximize the net profit associated with the inventory decision. In either case the validity of analysis will be affected by how accurately and completely we can describe and measure these relevant inventory costs. In general, costs are relevant in the inventory decision if they are controllable; that is, if the cost element will vary in a predictable manner based on the inventory decision, namely, the sizing and timing of the order quantities. Inventory costs are traditionally categorized into four basic types: purchase costs, ordering costs, holding costs, and shortage costs. We will discuss each in some depth.

Purchase Costs are simply the costs of the items that we hold as inventory. If we manufacture the item, this cost is the value of the item at that point in the system, to include all material and direct labor costs. For items that are purchased from outside the firm, this is usually the unit price we pay to our vendor. As an item moves through our logistics system, its purchase cost in the inventory analysis should reflect its fully *landed cost*, by which is meant the cost to acquire and move the item to that point in the system. Thus the landed cost would include transportation costs, tariffs, taxes, and any other costs incurred to make the item available at that point. The purchase cost is usually expressed and defined in an inventory model on a per unit basis, for example, as dollars per unit.

In some situations, the inventory purchase cost is essentially fixed with regard to the inventory decision and hence is not directly relevant to the inventory control decision. Suppose, for example, we carry inventory of an item to support a usage rate of 10,000 units per year. Further suppose that the item can be purchased at any time and in any quantity, but that the unit cost of the item will always be exactly \$5.00 per unit. In this case the average annual purchase cost will always be \$50,000, no matter how we operate the inventory, and hence the

purchase costs are not controllable and are not directly relevant to the inventory decision.

In many cases, however, discounts are available based on the timing or sizing of the inventory order. Sometimes a vendor will offer a specific discount for a very limited time period. More frequently, a vendor will offer a permanently available discount schedule based on the size of the order, with deeper discounts being offered for progressively higher order quantities. In many other cases, a consideration of the landed costs of an item will include this notion of a quantity discount because the nature of freight rates are such that large shipment sizes offer significant savings on a per unit basis versus small shipment sizes. Thus larger inventory orders will reduce the average per unit landed costs of the items in the inventory. Whenever these situations occur, it is important that the inventory logic carefully reflect the effect of the ordering policy on the total purchase costs.

Some items go through predictable seasonal pricing cycles which must also be taken into consideration in the inventory decision. In the extreme case, raw material prices may vary continuously as in the commodity markets. In these cases it is common to enter into futures contracts and hedging strategies to stabilize effective prices paid and to reduce risks. While important, these hedging procedures are beyond the scope of this section.

To make effective inventory decisions, we will need good estimates of the purchase costs for each SKU in the system. Most cost accounting systems can easily accommodate this level of analysis. Extending the estimate to include the fully landed cost is often more difficult because it requires associating and allocating average freight charges to the items.

Ordering Costs. In addition to the per unit purchase cost, there is usually an additional cost which is incurred whenever we order, reorder, or replenish the inventory. If we produce the item internally, there may be a *set-up cost* which accrues because we need to shut down a manufacturing line and "change over", reconfigure the line to make the specific item. For an item which is purchased from outside the firm, there is an administrative cost which is incurred in the purchasing function. This is the cost involved with processing the order, invoicing, paying the bill, auditing, and so forth. In many firms, the average administrative costs associated with placing an order can be surprisingly large. In inventory control, the usual convention is that the ordering cost occurs whenever we order and is fixed with regard to the size of the order. This is often a reasonable approximation. Production line set-up costs and times are usually unrelated to the size or length of the production run, and processing an invoice for an order of, say, 1000 units may cost no more nor less than processing an invoice for 10 units.

Estimates of ordering costs can also be derived from the firm's accounting records. Set-up costs are often available for individual items. Administrative ordering costs are usually expressed on a per order basis and are averaged across all items and all orders. As a first order approximation, for example, we could take the sum of all annual purchasing function costs (the payroll in the purchasing department, etc.) and divide by the number of purchases made per year to get an average cost per order.

Holding Cost. The costs that accrue due to the actual holding of inventory over a time period can vary considerably from one situation to the next. Many different kinds of costs can be considered as holding costs. The key characteristics of a holding cost are that the cost varies with the amount of inventory being held and with the length of time that the inventory is held. Holding costs are traditionally categorized as follows:

Storage Costs occur due to the costs of warehousing the inventory. Storage costs are usually a function of land and labor costs associated with running a warehouse, and thus can be expected to vary considerably from location to location. Storage costs at a location should vary in a nearly linear way across a broad range of storage volumes. For example, we might estimate that warehousing costs \$12.00 per square foot per year at a given site. Note that in analyzing the warehouse operating costs for inventory control purposes, we need to distinguish between the storage costs, which will vary with the amount of inventory held, and the throughput costs, which generally will not. Throughput costs include the costs to unload, inspect, put away, pick, pack, and load the property that moves through the storage facility. These costs may be considerable, but they do not generally vary with the amount of inventory being held or with the size of the replenishment order. They are determined simply by the volume flowing through the facility. It follows that warehousing throughput costs should be considered as a component of the landed cost of the item at the storage location.

Service Costs are out of pocket expenses that arise in addition to the physical storage costs. Two typical examples are insurance and taxes. Many firms will insure their inventories against fire and theft. The insurance rates will of course vary based on the nature of the goods and their location, but the basic rate will be applied against the amount of inventory and the length of time it is held. Many municipalities levy taxes on business inventories, and again the charge is incurred based on the average level or dollar value of inventory held over time.

Risk Costs are intended to recognize that there can be considerable financial risk associated with holding an item in inventory, because the item may become lost, stolen, damaged, or obsolete while it is being held. Some of these costs can be covered with insurance, but most firms treat damage and obsolescence as risks which are self-insured. Risk costs are significantly

different than storage and service costs in that they are inherently probabilistic in nature.

Capital Costs reflect the idea that an inventory ties up funds which the firm could be using for other productive purposes. As such, the inventory should bear an opportunity cost of capital so that financial decisions about the appropriate level of inventory in the firm will consider the costs of funding the investment in inventory. That being said, there can be differences of opinion about the appropriate capital cost to use in the analysis. Some would use the firm's marginal cost of new capital, while others would use the firm's weighted average cost of capital. Some analysts use the firm's investment hurdle rate in the inventory calculation, arguing that that the inventory decision is just such an investment analysis. The counter-argument here is that the hurdle rate is often intentionally set high to include a "risk premium" to accommodate the hazards and uncertainties of a new investment which can be very hard to quantify. Additional investments in inventories on an established line of business, on the other hand, should not be burdened with this additional risk premium.

The holding cost is usually estimated by accumulating all of these four components and averaging them across all of the items in the inventory system. Rather than expressing the holding cost on a per item or per unit basis, the cost is usually expressed per dollar value of inventory held. Thus a "30%" inventory cost would mean that the firm estimates that it would cost about thirty percent of the value of the item to hold the item as inventory for one year, including all storage, service, risk, and capital costs. At this rate, the firm believes that an inventory of finished goods with an average value of, say, \$10 million estimated at the cost of goods sold including all landed costs, would cost the firm an additional \$3 million per year simply to maintain the inventory.

Expressing the holding costs as a percentage of the value of the item averaged across all items is simple and intuitive, but this level of aggregation can be very inappropriate. Capital costs, insurance, and taxes will clearly vary with the dollar value of the item, and so these elements will be well approximated. Physical storage, damage, and obsolescence costs are far more item specific. Applying a system-wide average holding cost to each of the items in the inventory may result in serious distortions of the true holding costs for a given item. In these situations, holding costs can and should be estimated for individual items or selected classes of items.

Another potential component of the holding cost that is sometimes discussed is the **masking cost**. The idea here is that by holding inventory, the firm will cover up or "mask" operational problems, and the costs of these problems should therefore be somehow charged to the true total costs of the inventory. This idea has been widely discussed in the framework of the Just-In-Time inventory approach. In a manufacturing operation, for example, the firm may be buying a component part which has a high defect rate. If the firm carries

a liberal inventory of the component part, then the assembly workers will simply discard the defective units as they are found, and the problem will be ignored. The firm will be wasting money on defective components, and worse still, occasionally a defective part will slip through into the assembly, resulting in quality control problems, rework, defective finished goods, unhappy customers, etc. On the other hand, if there had been no additional inventory, that is, if the firm carried only enough inventory to satisfy its most immediate requirements, then each defective component would have created an immediate, obvious problem, and management would have focused its attention on the solving the supplier's quality problem. In this view, the costs of the defective components, the rework, the customer ill will, and all the rest, are masked costs which should be included as part of the costs of holding the inventory.

This is an interesting concept which has proven very difficult to implement. Part of the problem is that these costs are highly probabilistic. What is the probability that a given inventory will generate a masked cost, and what will be the magnitude of that cost? A more serious problem with the concept is the notion that the masked costs are somehow "caused" by the inventory in the same sense that storage costs are generated by inventory. For example, in the example of the defective component, was the inventory really the root cause, or were all of the problems really caused by the management that was ignoring the scrap rate on the component? Assigning the blame to the inventory is a great deal like blaming the high rate of automotive collisions on mandatory auto insurance laws rather than on bad driving.

Shortage Costs are the fourth major type of costs which are usually attributable to the firm's inventory control policy. Purchase, ordering and holding costs can be thought of as the costs of having inventory, while shortage costs result from not having inventory, or from not having enough inventory at the right place at the right time. When a demand arises which cannot be satisfied from available inventory an inventory shortage occurs. The cost resulting from this shortage will depend on the specifics of the situation. Many situations can be envisioned, but the four most common scenarios are as follows:

Backorders occur when the customer is willing to wait for inventory to be made available. In business to business transactions, both industrial and commercial, backordering in an out-of-stock situation is a common practice. Costs accrue because the firm must maintain a system to record the backorder, maintain it, and fill it when stock becomes available. There may be additional costs due to shipping. For example, in filling an order from stock the firm may routinely charge the shipping costs to the customer or they might be included in the price charged. On a backorder, these costs may be waived, or the firm may use expedited, premium freight services to make up for the delay, and these additional costs would not be passed on to the customer. In commercial practice, an order will frequently contain a number of different items. If all are in

stock, one large, relatively low cost shipment will result. If an item in the order is out of stock, the firm will usually ship the in-stock items immediately, as a slightly smaller, less cost effective shipment. As the out-of-stock items become available, each results in a small, relatively expensive shipment, particularly if expedited freight is involved. As a consequence, the total shipping costs of the entire order can be considerably higher as a result of the out-of-stock items. Backordering can be seen to result in incremental order processing, handling, packaging, and shipping costs which can be substantial. But the most important component of the backorder cost may be the "ill will" or implicit penalty cost which recognizes that the customer does not want a backorder. The customer may eventually retaliate by reducing the number of orders, demanding compensation, or by finding a new source. This penalty cost component can be very probabilistic, subjective, and difficult to estimate, but it should not be ignored. Once the situation has been analyzed and the component costs have been estimated, the backorder cost can be thought of as arising per backordering incident, per unit backordered, or per unit backordered per period of time the backorder is allowed to exist.

Lost Sales occur when the customer responds to an out-of-stock situation by canceling the demand. This is the normal consumer practice at the level of the retail shelf, but it also occurs in commercial practice when orders are handled on a "fill or kill" basis. When a lost sale occurs, the firm loses the gross profit margin on the unit which could have been sold. In many cases, the loss extends far beyond the margin on the item which is actually out of stock. In a retail convenience store, for example, some items like soda and beer will have relatively low gross margins, while other items like pretzels and other snack food may have much higher margins. If I stop here for beer and pretzels, and the store is out of beer, will I buy the pretzels here, and look elsewhere for the beer? Almost certainly not; I will generally just leave and buy both items at the next store. Thus the lost sale cost associated with a "magnet" item, like the beer in this example, should include the lost margin on the cluster of item sales which are likely to be lost when the "magnet" item is out of stock. As you can see, accurately estimating the totality of lost sales costs is not necessarily simple.

In addition to the lost profit margin, there is a still more difficult to estimate "ill will" component which should be included to recognize that some customers will eventually "punish" the firm for poor inventory availability. Many firms are uncomfortable with the inherently subjective nature of this "penalty" component and therefore ignore it. This will clearly result in an underestimate of the "true" cost of a lost sale. If these costs are used in the inventory analysis, lost sale costs are usually thought of as arising per unit of demand lost.

Lost Customer Costs can arise from out-of-stock situations as well. While it may be very difficult to assign any specific "customer defection" to a single out-of-stock situation, there is clearly a causal linkage between poor inventory availability performance and the retention of loyal customers. For

example, careful market research might suggest that, on average, every fifty lost sale units of an item might result in one lost customer; that is, that a customer will respond to this level of inventory out-of-stock experience by taking all future business to another firm. If we were to believe that this was the case, what cost would we assign to this outcome? This lost customer cost can be out of all proportion to the cost or profit margin on the item which is stocked out. From this perspective, we can think of the cost of the lost customer as the net present value of the stream of lost gross margins of all the future purchases of this item which now will not occur. In addition, we lose the margins on all the future sales of other items which this customer now will not buy from us. And there is also the impact of all the negative word of mouth advertising which will naturally result. This kind of cost analysis is inherently "soft" and subjective, but it is well worth considering.

Another way to think about the cost of a lost customer is to consider the "replacement cost". For example, consider the firm's total annual marketing budget. What percentage of it is intended to generate new customers, as opposed to retaining current customers, increasing their purchases, etc.? Now, how many net new customers were generated? These two numbers suggest what the average cost to generate a new customer, or to replace a lost customer, really is. Most firms have no idea what this number looks like. If we don't what a lost customer is worth, how can we rationally decide how much we should be willing to invest to hold onto one?

Disruption Costs occur in a production environment when a shortage of raw materials or component parts causes an interruption to the planned activity. These costs are frequently out of all proportion to the value or cost of the goods in question. For example, a tire manufacturer provides complete tire and wheel assemblies to a truck manufacturer on a just-in-time basis. The tire manufacturer delivers "two hour's worth" of wheels directly to the assembly line every two hours. If the tire delivery is late, the entire assembly line must be shut down because the trucks cannot be driven off the line without wheels. The disruption cost here would be extremely high. In fact, the truck manufacturer shares this risk with the tire manufacturer. Under the terms of their purchasing agreement, the tire manufacturer pays a substantial fine, on the order of tens of thousands of dollars per minute, whenever the line must be shut down due to the unavailability of wheels. In this case, the disruption cost is very explicit. In every case where the lack of a raw material or component would cause a disruption or delay, we should establish an estimate of the true cost of the disruption. This cost can then be used to determine the appropriate amount of safety stock or lead time padding needed to mitigate the risk of the disruption.

A Final Word on Estimating Costs. In cost-based inventory analysis, questions frequently arise as to which costs should be included and which should

be left out, usually on the basis that they are somehow "fixed" or "sunk" costs. The general rule is that a cost element should be included:

- A. If it is material; that is, if it is a relatively substantial cost, and
- B. If it is controllable; that is, if the level of total cost will change in a predictable way based on the inventory decision or logic.

It will be useful here to comment on some cost estimating situations which often arise:

Storage Costs. In analyzing the inventory position for a specific item, some will argue that storage costs are irrelevant because the decision on one single item in our large system will not result in the opening or closing of a warehouse, hiring or firing employees, etc. Thus the storage costs are essentially fixed with regard to this item and should be excluded from the analysis. Apparently, the marginal storage costs of this item are zero. There is some merit to this line of thought, but the decision should be carefully thought through. If we were to apply this rule to every item, we would never consider storage costs. But suppose we were to perform a system-wide analysis which resulted in a permanent, fifty percent reduction in system-wide inventory levels. We would certainly reduce our storage capability (and hence the storage costs) proportionately, even if it took a while to accomplish. In most analyses, it is most logical to take this long run, system-wide perspective, which is to include all costs which are controllable in the long run.

Ordering Costs. This same line of argument is often applied to ordering costs as well. After all, a few more orders, one way or the other, won't really affect the annual budget of the purchasing department. On the other hand, if we were able to reduce the number of purchase orders by say, ninety percent, wouldn't we take the savings that results? Recognize also that if we consider the marginal cost of purchase orders to be zero, the inventory algorithm will quite properly begin to overwhelm the purchasing system with large numbers of small orders. Here again the better approach is to estimate the true marginal cost of placing an order and to include the cost in the analysis.

Shortage Costs are frequently ignored because they are often not "out of pocket" and because their estimation is thought to be highly subjective and speculative. Management can be very uncomfortable in using these figures to drive a multi-million dollar investment decision. However, it is very difficult to determine the appropriate dollar investment in safety stock if you cannot specify the dollar loss you are "insuring" against. As an alternative, firms will rely on non-cost based safety stock procedures, such as target service levels or fill rates, which can be used to size the safety stock. It is useful to realize that any such service-based stockage policy implies a specific shortage cost. That is, for every possible choice of inventory service level there is an implied or imputed shortage

cost such that the given service level would in fact be optimal if that implied shortage cost were in fact true. If the firm uses such service-based logic, it should certainly also calculate these imputed costs and assess their reasonableness. This approach will be developed further in the section on safety stock which follows below.