1

Supply Chain Management

1.1 WHAT DO WE MEAN BY LOGISTICS?

Logistics has quite a long history, whose origins predate by far the initial attempts to make it "scientific." Many engineering schools were born because of the need for building better military fortifications and weapons. Logistics followed a pattern common to that of many fields in engineering: Military applications gave an important impulse to its development. While relatively small armies in the past could sustain themselves also by robbing local populations, proper management of supplies was required at later times to support larger armies in need for ammunition and a significant amount of food. Napoleon, who is acknowledged with the motto "An army marches on its stomach," is considered an innovator in this respect, because (what we now call) supply chain management afforded his armies a far greater degree of mobility than his rivals. Logistics has played an increasing role in later conflicts, like the American Civil War (ACW), where transporting supplies and troops was accomplished by an array of transportation means including supply wagons, rail, ships, and (in the Western Theater) rivers. The role of logistics can be appreciated by considering how the availability of supplies is of no use if the supplies cannot be routed to destination, whereas clever organization may make good enough use of scarce resources. A paradox in Confederate logistics during ACW was that an economy strong in agriculture and weak in industrial power, compared to its Union counterpart, succeeded in maintaining a flow of weapons and ammunitions, whereas troops often starved because of lack of food.[1] Indeed, some military academics are reported to say that "amateurs study tactics, professionals study logistics."

Military applications continued to play a prominent role in the development of scientific logistics in the 20th century.[2] The quantitative approach to management problems is typically associated with Operations Research, whose origin can be attributed in part to the need of managing the supply chain across the Atlantic Ocean during World War II.[3] However, we should not think that the scientific approach to logistics is that recent. For instance, the well-known Economic Order Quantity (EOQ) formula for inventory management dates back to the early 20th century, since it was published in 1913[4]; furthermore, the manifesto of Taylorism[5] was published in 1911, but its roots can be traced back to a rationalization process in manufacturing, which had been quite active during the 19th century.

Given this long history, we should not be surprised that the term "Logistics" has now a rather wide and often ambiguous meaning. Indeed, several professional and academic organizations have attempted to draw the line, pointing out what we should mean by this term. The U.S. Council of Logistics Management proposed the following definition:

> *Business logistics is the term describing the integration of two or more activities for the purpose of planning, implementing and controlling the efficient flow of raw materials, in-process inventory and finished goods from the point of origin to point of consumption. These activities may include, but are not limited to customer service, demand forecasting, distribution communications, inventory control, material handling, order processing, parts and service support, plant and warehouse site selection, procurement, packaging, return goods handling, salvage and scrap disposal, traffic and transportation and warehousing and storage.*

The term *business logistics* emphasizes a separation from other fields, such as urban transportation, which could be included in a more general notion of logistics. The definition we have reported is not very recent, as it dates back to 1979, but it includes both management issues and material handling issues, which are more physical in nature. This book is only concerned with management issues, not with physical activities which might be labeled as

[1]See: R.K. Krick, The Power of the Land, in: A. Sheehan-Dean (editor), *Struggle for a Vast Future: the American Civil War*, Osprey Publishing, Oxford, 2006.

[2]Those of us who are sane enough not to appreciate the grim arts of war too much, may find some consolation in thinking that the same approaches can be used to route huge amounts of essential supplies, in a short time span, to areas struck by natural disasters.

[3]Another element in the birth of Operations Research was queuing theory, initially developed to model telephone traffic. It is worth remembering that the celebrated simplex method to solve linear programming problems was developed in 1947 by George Dantzig, who worked for U.S. Air Force.

[4]See: F.W. Harris. How many parts to make at once. *Factory: the Magazine of Management*, Vol. **10**, 1913, pp. 135–136. Reprinted in *Operations Research*, 1990, Vol. **38**, pp. 947–950.

[5]F.W. Taylor. *The Principles of Scientific Management*. Harper & Row, New York, 1911.

"industrial" logistics. This is certainly *not* to say that industrial logistics has a lesser role, or that there is no interconnection between hardware and managerial issues. Some management activities have no sense if the underlying physical process is not properly designed and if certain technologies are not exploited. Our aim is to define a consistent and relatively limited scope, in order to offer a pedagogical treatment of selected material at a suitably deep level, rather than offering a superficial handbook covering all possible topics. As we stress below, solid foundations are essential to any practitioner, as general principles have to be twisted and adapted to many diverse and peculiar settings, and a superficial listing of cookbook recipes is actually of little use, if not counterproductive in case these recipes are applied improperly.

Apparently, the definition above includes too many things. However, modern integration trends have given rise to Supply Chain Management (SCM) as an almost all-encompassing discipline. On the supply side of the chain, increasing emphasis is given to supplier relationships management, purchasing, and contract design. On the other end of the spectrum, customer relationships management (CRM) is another example of an issue which is gaining relevance. Information Technology (IT) had a dramatic impact too, thanks to the rise of Internet, which made electronic commerce, online auctions for products and services, and the sharing of large databases possible. As far as information systems are concerned, the introduction of Enterprise Resource Planning (ERP) systems has made the case for the interconnection with other functional areas, such as manufacturing,[6] accounting, etc. And if this does not look confusing enough, the list of complications could go on and include other factors:

- The reduced lifespan of products and the need for customization imply that the supply chain has to be continuously redesigned. Even product design may interact with logistics. For instance, design for supply chain management has been successfully applied by Hewlett-Packard.[7]

- Globalization has introduced a new array of risk factors which impact SCM, such has exchange rate risk and, at a higher level, political risk.

- The availability of several transportation modes and the concentration of production into large sites have a deep impact on transportation management.

[6]Indeed, in many practical settings, we cannot deal with distribution logistics without paying due attention to production. From a methodological point of view, many models and modeling techniques we illustrate in the book are often included in books on manufacturing management.

[7]See: H.L. Lee, C. Billington, and B. Carter, 1993, Hewlett-Packard Gains Control of Inventory and Service through Design for Localization. **Interfaces**, Vol. 23, pp. 1–11.

- Revenue and yield management[8] have a prominent role in the air transportation and in the service industry, but they are likely to see an increased role in distribution too (think of price cuts at the end of the selling season in many retail chains).
- Environmental issues dictate that we also pay due attention to *reverse* logistics.

All of the above, and more, has something to do with Supply Chain Management. Trying to cover such a wide spectrum or topics and issues in one book is a hopeless endeavor, unless one is willing to just compile a list of buzzwords. We believe that students (and practitioners) should have a firm grasp of basic principles of distribution logistics. Armed with a solid background, they can tackle new developments with confidence. Quantitative models and methods play a fundamental role in developing basic skills, and indeed this book is more quantitative oriented than others in this area. However, we did not aim at writing a high-level research survey for Ph.D. students. We only outline problems and solutions, using both toy examples to build intuition and real cases when appropriate. Moreover, we should never forget that quantitative models may be implemented in a computer program, but they are ultimately applied by people. People have incentives, possibly unwritten ones; this applies both to single individuals and to organizations. Indeed, distribution logistics typically crosses borders between organizations, and understanding incentives and organizational barriers is a prerequisite to successfully apply any "scientific" solution.

1.1.1 Plan of the chapter

After insisting on what we do *not* include in the book, we would better explain what we *do* include. This chapter lays down the foundations for the next ones, according to the following plan.

- A distribution network is characterized by a physical arrangement of facilities, such as warehouses and transit points, on a possibly wide geographical area. In section 1.2 we illustrate typical structures of distribution networks. The physical arrangement of facilities does not tell the whole story, as goods flow in the network by some transportation means (e.g., trucks or rail). Inventory and transportation management strategies contribute to the definition of a distribution network. Furthermore, information flows must be described too.
- When designing a distribution network, we should make our decisions in a way that supports a specific strategy. There is no single "one-

[8]Revenue and yield management are essentially dynamic pricing policies. They have a prominent role in the case of goods which cannot be stored, such as seats on an aircraft; transportation services can also be priced dynamically, as well as perishable items.

best-way" strategy that works in all possible settings. A strategy is a compromise between the need of achieving a good competitive position, according to a selected profile, and the need of keeping costs low. Competitive factors, cost drivers, and possible strategies are outlined in section 1.3.

- A distribution network typically includes locations in which goods are stocked. Common wisdom maintains that inventories are the source of a long array of evils and should be kept as low as possible. In fact, inventories are a source of many relevant costs, but they play specific roles in achieving a certain competitive position. Hence, they must be properly managed and we should have their functions very clear in mind. Section 1.4 illustrates the roles of inventories.

- A recurring theme in this book is uncertainty. Demand uncertainty is the single most relevant complicating factor in distribution logistics. Good forecasting procedures may be used to predict future demand, but they can only reduce rather than eliminate uncertainty. Even if uncertainty cannot be eliminated, it can be managed. In section 1.5 we start outlining a few ways to deal with uncertainty.

- Goods move on a distribution network, from factories in which they are produced, through warehouses and transit points, to retail stores. Managing transportation is another relevant piece in the overall puzzle. Section 1.6 illustrates some basic ways to define a transportation strategy.

- The flow of goods is what is typically associated to logistics, but the flow of information is just as important. Any decision procedure is based on some piece of information, but without information sharing, certain procedures are simply not feasible. Information sharing may be difficult in a large firm consisting of several branches, let alone a supply chain involving different firms. Furthermore, assigning decision rights in a supply chain involving several actors is not a trivial task. Section 1.7 outlines a few issues related to information, incentives, and decisions.

- The structure of a network is something that should not change too quickly, since the decision to build a facility may be made considering a relatively long time horizon, say years. A recent tendency is to lease warehouses, which contributes to shorten the time span of these decisions.[9] Nevertheless, moving all the goods from an old warehouse to a new one is not something we want to do on a monthly basis. On the contrary, a change in the inventory management strategy can be

[9]Another factor which calls for frequent changes in the supply chain is the reduced life-cycle of products.

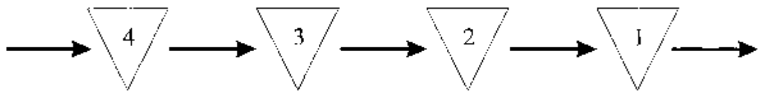

Fig. 1.1 Linear logistic structure.

achieved on a shorter time span, and transportation must be managed daily. Hence, different decisions may have different time horizons and pertain to various hierarchical levels. In section 1.8 we introduce strategic, tactical, and operational decisions. These should be regarded as loose guidelines, since sometimes it is hard to draw the line between the levels, due to tight interactions between different them.

- There are some recurring expressions in Distribution Logistics, and more generally in Operations Management, such as make-to-stock, make-to-order, push, and pull. They have raised quite a bit of controversy, as sometimes they are used ambiguously. Indeed, they do not really define specific strategies, but they do define *attributes* of possibly hybrid decision strategies. In section 1.9 we illustrate the meaning of these terms as features of decision strategies.

- Last but not least, to tackle all of the above problems we may take advantage of models and methods. Quantitative approaches play a prominent role in the book, which is not to say that they should be applied with a blind faith in their power. Section 1.10 helps in classifying quantitative models, including those which are quite useful but are not dealt with here; the most notable example is discrete event simulation.

1.2 STRUCTURE OF PRODUCTION/DISTRIBUTION NETWORKS

From a physical point of view, a supply chain consists of possibly several stages where items are produced, transformed, assembled, packaged, and distributed to consumers. The simplest structure is illustrated in figure 1.1, where we see a linear arrangement of nodes. Each node in this chain can be more or less complex. The first node is likely to be a factory, where items are produced; we deal with this node as a black box, but a manufacturing system would consist in turn of several machines, laid out according to a certain pattern. From our distribution point of view, these details are not quite relevant per se. However, the arrangement of the manufacturing system has a definite impact on performance measures such as flow time, i.e., the time that an order takes to go through all of the stages required by its technological cycle. The manufacturing flow time is clearly relevant from the supply chain point of view. Thus, we do not investigate the internal structure of the nodes and treat them as black boxes. However, the performance (cost, lead time, etc.) of each

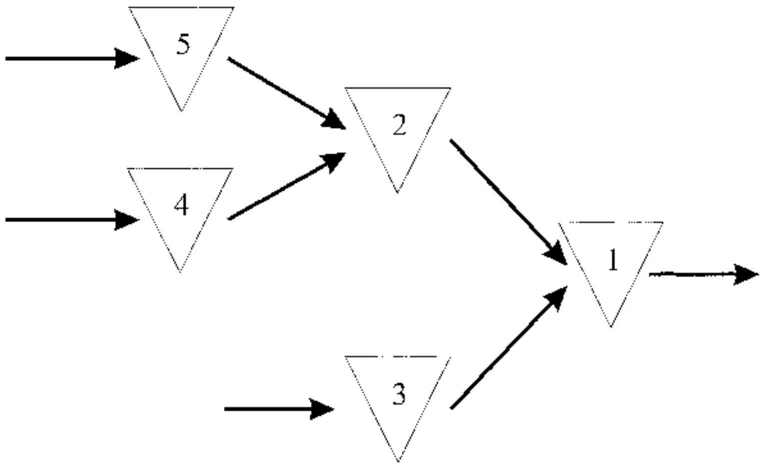

Fig. 1.2 Supply chain structure with assemblies.

black box is very relevant to us. The network could be extended to the left, and include the production of raw materials, but any analysis has to focus on a portion of the overall chain. Proceeding to the right in the figure, we may find other stages at which material is transformed; we should pay attention to the increase in value of the product, which affects the overall economic performance of the network. After the whole chain of transformations, the products may flow through other stages, at which material is simply stocked in a warehouse, until the retail store is reached. Factories may have inventories too, both inbound and outbound.

Along a linear chain, we may have transformations and transportations of items. However, assembly of components into end items is a common occurrence. When items from different sources are assembled, we get a converging structure like that illustrated in figure 1.2. Readers with a manufacturing background could be tempted to interpret the convergent network in the figure like a bill of materials, i.e., a technological representation of how an end item is obtained by assembling components and possibly complex subassemblies. Actually, what we are representing here is the geographical structure of the network, where components can be produced in a continent and assembled in another one. In a convergent network, we clearly see the need for proper synchronization in the material flow: If we miss even one, possibly low-cost component, we cannot assemble the product we need.

Finally, figure 1.3 illustrates an arborescent (or divergent) network which is typical of pure distribution. Here node 1 could be a large warehouse located near a factory producing an item, nodes 2 and 3 might be regional warehouses, and the remaining nodes could be retail stores (in a real network, there would be much more retail stores than depicted). In a pure distribution

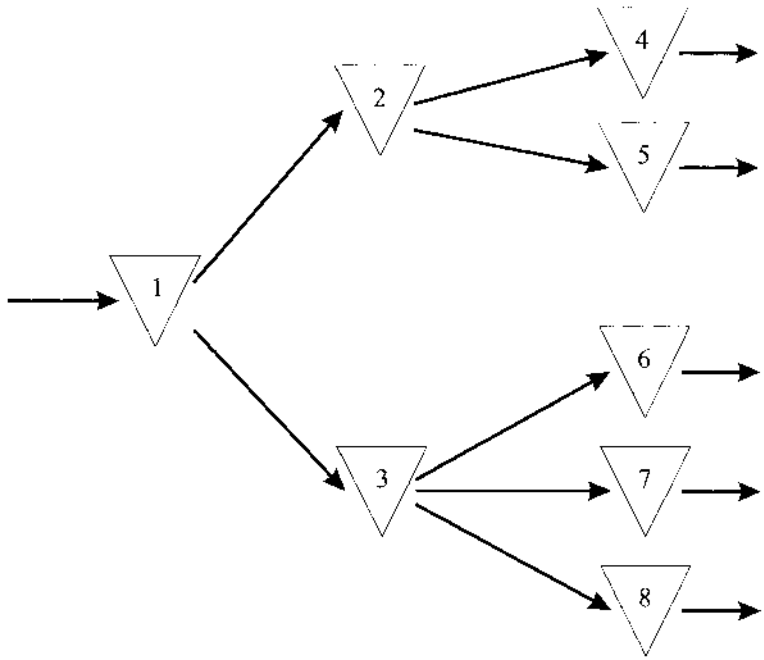

Fig. 1.3 Pure distribution (arborescent) network.

network, the product is always the same.[10] However, whenever material is transferred downstream, we commit it to a certain section of the network. Such an allocation decision is absent in the previous cases, and it must be made with care when material availability is scarce. One could wonder why intermediate stages are needed; after all, they are a cost. We will consider the roles of intermediate stages in depth in chapter 2. Intermediate nodes might help the company in exploiting economies of scale in transportation and/or to reduce the impact of demand uncertainty. We should note that intermediate nodes can be distribution warehouses, but they can be also simple transit points with no facility to store inventory; alternative terms for the last case are "transshipment nodes" or "cross-docking platforms."

The three structures we have illustrated are just basic prototypes. A real-life supply chain is a hybrid of all of them, with many variations. For instance, in the distribution network of figure 1.3, material flows downstream according to a regular pattern, stage by stage. In practice, some retail stores could be served directly from node 1. We will see that this depends on the demand

[10]The lack of physical transformations does not imply that the cost of items does not change; as an example, consider customs duties we may have to pay when crossing certain borders.

volume; when this is large enough, we do not need intermediate nodes to take advantage of economies of scale in transportation. For example, we might be in a position to fill a full truck leaving the warehouse to visit a given store. Another variation, with respect to scholastic cases, is the reverse flow of materials. In the previous figures, we see material flowing downstream, but recycling and the need to collect waste call for proper management of *reverse* logistics. The increasing concerns for the environment make such issues more and more relevant. Finally, we may have flows of materials between peer nodes, i.e., stages which are located at the same level in the network. These *lateral* shipments can be used to reallocate material among stores of large retail chains in case one is experiencing a stockout and another is overstocked.

A network design problem calls for structuring a possibly large supply chain, locating facilities, deciding their capacity, and optimizing the transportation of material among them. This is a very difficult task, as we shall see in chapter 2. Fortunately, we are often interested in the *partial* redesign of a network, which makes the task considerably easier. However, the shorter and shorter life cycle of products calls for the continuous redesign of supply chains.[11]

1.3 COMPETITION FACTORS, COST DRIVERS, AND STRATEGY

When managing a supply chain, the natural aim is providing the customer with a suitably good service, and doing so at a suitably low cost. By "good service," we mean that the customer should get *what* she wants, *when* she wants it, and *how* she wants it. Other factors could be relevant, such as after-sales service, but even if we focus on the minimal set of attributes that make a good service, we see that there is no single dominant strategy: There is no possibility of being first in class along all possible dimensions, at a reasonable cost. What we need is a clear view of the dimensions on which we compete, in order to get priorities straight. In the following sections we illustrate a few examples of attributes which define competition factors; then we list a few sources of cost that we must keep under control; finally, we illustrate how all of the relevant dimensions can be traded off, by prioritizing competition factors to define a strategy.

1.3.1 Competition factors

Say that you are a customer wishing to buy a certain good. What are the attributes that are important to you? Probably, most people would point out **quality** requirements. Of course, quality of the goods is key factor influencing

[11] Hewlett Packard provides an excellent example of using software tools based on quantitative methods to design supply chains in a dynamic environment; see [2].

consumers' choice, but it is itself a complex concept encompassing multiple dimensions. The quality of a car can be measured through the number of safety features, the top speed, gas consumption, acceleration, etc. Moreover, quality can be measured through the *target quality* (i.e., the quality the product should have, according to its design) and *conformance quality* (i.e., the ability of the single item to meet the target quality over time). Also, the quality of the good could be traded off against price, depending on which market segment we want to address. Moreover, quality is relevant not only in terms of goods, but also in terms of *service*. Indeed, there are complementary services which may contribute to establish a reputation. Consumers can return the merchandise they bought to many mail retailers (as well as to brick and mortar retailers in countries such as the U.S.A.). Other services are more and more relevant in times of increasing environmental concerns; we have already mentioned the role of reverse logistics and the possibility of returning packaging materials, used products, etc., which contribute to the positive image of an environmentally responsible supplier. After-sales services are specifically important for durable goods whereas installation support is very important for complex systems such as high-end audio and video systems.

If we think of distribution services *per se*, fast delivery may be important, but *dependability* may be even more. So, waiting for a long **Delivery Lead Time** (DLT) may be unpleasing, but a very uncertain and unreliable DLT may be even more annoying. In fact the possibility of tracking shipments or to check order status, possibly via Web, is typically offered by couriers, such as DHL and Fed Ex, by Internet-based sellers, and service centers of non-durable goods. From the consumer's point of view, DLT must be zero for some products: No one would like to wait a few days for a bottle of milk. However, the DLT for milk is not zero from the point of view of the retail store or of other actors along the supply chain. Yet, we will see that a zero DLT may make the management of inventories much easier. On the one hand a non-zero DLT provides us with some advance information that can help us improve performance (e.g., reduce inventories or increase service level). On the other hand, exploiting this information is all but trivial and complicates modeling substantially.

At the other end of the spectrum, engineered-to-order items have a long DLT: No one would expect to find a radar system on the shelves. In between these extreme cases, there is an array of intermediate possibilities. DLT is linked to the structure of the network, the transportation means adopted, and the inventory levels and their deployment in the network. If large amounts of goods are held near the customers (say at the stores), DLT is short; it can also be reduced if quick but costly transportation services are used. So, we see that there is a tradeoff between DLT and different types of cost.

Example 1.1 CHL is an Italian retail chain of information technology products. An important feature of its strategy is that it does not maintain inventories at retail stores, which are just used to collect orders and to deliver items to customers. This results in a significant reduction in inventory levels, which is particularly relevant for items characterized by very fast obsolescence and thus high cost of inventories. □

Another relevant competitive weapon is **assortment**, i.e., the variety of products offered. For a manufacturer, this means offering a large catalogue and the possibility of customizing an end item according to customers' wishes. For a retail store, this means offering a large set of alternative items on its shelves. In both cases, we see that variety comes at a cost. Also, we can trade off assortment with DLT. If products are customized to order, we need some time for this operation and customers shall be willing to wait. If you offer a large assortment with zero DLT, you have to keep a lot of items in inventory, each one with a possibly low and hardly predictable demand. However, variety may be an important and valuable asset to attract customers. Indeed, there may be a positive feedback, when variety increases demand, thereby easing some of the difficulties associated with low levels of demand.

Another relevant feature of the supply chain is the **flexibility**, that is the ability to adapt to changes and exceptional conditions. For example, a flexible supply chain can fulfill an extremely important order in an exceptionally short time. We can have different kinds of flexibility according to the variable that raises the need for a change. We call *product flexibility* the ability to adapt the product to customers' needs. For example, the ability to configure the product to customer specifications might be crucial for complex products such as furniture or cars. A company that carries inventories of components and assembles them to order usually can achieve a great deal of flexibility with limited resources (provided customers are willing to wait while components are being assembled). Think of the large number of different sandwiches one can prepare with just a few basic components! We call *flexibility to product innovations* the ability to manage the introduction of a new product. To achieve this kind of flexibility the company might need to buy flexible production systems and might want to carry components over, that is use components and subsystems from previous generations of the product. Such kind of flexibility is more and more important nowadays given the growing importance of new products and product novelty. We call *delivery flexibility* the ability to adapt deliveries to customers' needs. For example, the ability to deliver rush orders or manage luggage of VIP clients with a tight connection in a hub-airport might be crucial. We call *volume flexibility* the ability to increase/decrease production and distribution quantities on a short notice. This flexibility is expecially valued in markets with a sharply seasonal pattern, such as Christmas gifts, etc. This flexibility can be gained through both spare resources (e.g., spare capacity), flexible resources (e.g., temporary workers), or appropriate planning (e.g., we might produce/distribute all products with a predictable demand before the peak of the season so that during the peak we can use the limited production/distribution capacity to manage just the uncertain part of demand).

1.3.2 Cost drivers

Keeping costs under control is a fundamental factor in supply chain competition. We should state quite clearly that cost minimization per se need not be a winning strategy; a strategy is a good tradeoff between the objective of minimizing costs and the objective of maximizing other competitive performance metrics such as quality, delivery, service, etc. Keeping this in mind, we should list the typical cost drivers in supply chain management, in order to set the stage for decision-making approaches. Before doing so, we should classify costs according to a couple of dimensions.

- *Costs can be linear or nonlinear.* Consider an arbitrary activity (e.g., how many parts we make or buy), and denote its level by a decision variable by x. A linear cost function is something like $f(x) = cx$, where c is a unit cost. More generally, if we have N activities indexed by i, a linear cost function has the form $f(\mathbf{x}) = \sum_{i=1}^{N} c_i x_i$; note how linearity implies that costs pertaining to different activities are simply added. Otherwise, we have to deal with a nonlinear (possibly discontinuous) cost function. Examples of nonlinear cost functions are $f(x) = x^{0.6}$ or $f(x_1, x_2) = x_1 x_2$. Consider, for instance, purchasing large amounts of some component; a discount might be offered if the purchased quantity is above a given threshold. In such a situation, we have an economy of scale; diseconomy of scales occur when scaling an activity level up increases the related cost more than proportionally. Interactions among activities may also result in a nonlinear total cost function.

  In practice, costs are always nonlinear, but sometimes they can be suitably approximated by linear functions, at least for small variations of the level of activity (say number of units purchased or produced). When formulating an optimization model (see appendix B), keeping everything linear is an important concern in order to limit the computational effort required for solving the model. Even when assuming a linear cost function is too far from reality, nonlinear costs can be approximated by piecewise linear functions (see section 2.3) whereas in the general case they can be fairly different.

  We may also recall two important concepts. Consider a generic cost function $c(x)$. The value of the first-order derivative $c'(x)$ is called **marginal** cost. The marginal cost is constant for a linear cost function, but not in general. The **average** cost is $c(x)/x$; we may see that average and marginal cost are the same for a linear cost function.

- *Costs can be fixed or variable.* In accounting, a cost is **fixed** if there is nothing we can do about it in the short term.[12] For instance, the

[12] Strictly speaking, accounting professionals use "period" and "product" costs.

cost of a plant is fixed from the point of view of short-term operations (consider, e.g., rent, depreciation, or cost of fixed personnel). The direct production cost is **variable,** since we can change it through production decisions on a much shorter time scale. Of course, in the long run all costs are variable, so the distinction is a matter of time scale. Nevertheless, such fixed costs do not (or at least should not) influence current decisions; they may contribute a constant term to an objective function in an optimization problem, but this does not change the optimal solution. In the short run, these fixed costs are constant, no matter what the short term decisions are. So in a way they are simply irrelevant for decisions making processes. Sometimes, the term *sunk* cost is used to refer to a cost which has been paid and no future decision has any influence on it.

In this book, we will use fixed/variable costs with a slightly different meaning. If a cost function can be expressed by

$$c(x) = \begin{cases} F + cx & \text{if } x > 0, \\ 0 & \text{otherwise,} \end{cases}$$

we refer to F is the fixed cost. Hence, what we mean by "fixed cost" is a cost that does not depend on the value of a decision variable, provided it is strictly positive.[13] The typical example of fixed cost in this vein is a fixed ordering cost, i.e., a cost that we pay whenever we order, whatever amount we order. Clearly, such costs might encourage ordering larger quantities, resulting in economies of scale.[14] Hence, fixed cost in this sense do influence decisions, unlike fixed costs in the accounting sense (for the sake of clarity in the remainder of this book we will call these sunk costs).

Fixed costs may result in piecewise constant cost functions. Consider the cost of transporting an amount x of some good, and assume that there is a fixed cost component, that we pay for each truck we use. Depending on x, we may have to use one truck or two. This induces a discontinuity

[13] Sometimes, the term fixed *charge* is used to avoid ambiguity.

[14] Notice that in Economics the term "economies of scale" has a slightly different meaning, since they are regarded as a long term phenomenon. When we face economies of scale, the long term average cost decreases as the production volume (per unit of time, say per year) increases. When economists say "it is a long term effect," they really mean that we can observe such a reduction in the average cost when we compare different plants (or, more generally, infrastructures) with different capacities. On the contrary, the effect of the production volume on the costs of a given plant is a short term effect. As such, economists do not consider this to be related to economies of scale. In this book we use the term economics of scale in a broader sense. Therefore, in this book the economics of scale lead to the reduction of the average unit price and might be due to the dilution of some fixed costs when the level of activity (say the production volume, the purchase quantity, or any other relevant level of activity) increases. We disregard the distinction between short term decisions and long term ones.

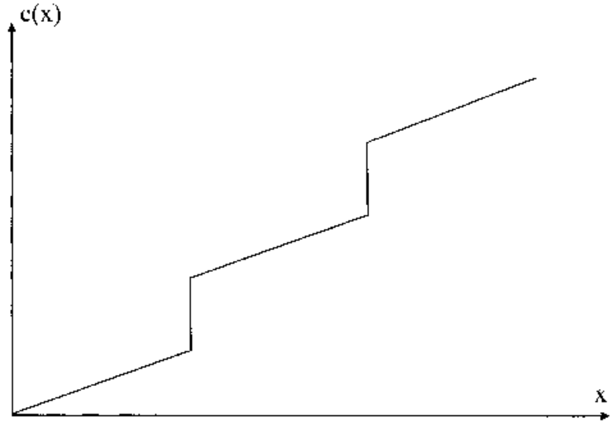

Fig. 1.4 Semivariable costs.

in the cost function, which might include a piecewise constant term. Sometimes, the term **semi-variable** cost is used to refer to such a case (see figure 1.4).

We stress again that we cannot really draw a thick line between the concepts above: A linear cost function can be a suitable approximation of a nonlinear one, and a fixed cost may be transformed (at least partially) into a variable one by suitable arrangements. So, we should just consider the above classifications as useful guidelines, which are best illustrated by a few examples.

We have seen that a supply chain is, from a physical point of view, a network of facilities on which goods are stocked and transported. A first set of costs is associated with building and maintaining facilities. These costs are sunk when we are operating the network, but they are a result of a decision when we are *designing* the network. The cost of a facility is a possibly complex function of its type, location, and capacity. A pure transit point is typically less expensive than a distribution warehouse. We need to find a suitable approximation of the cost associated with building and operating a facility, and this is certainly not a simple linear function. Some costs are fixed, such as those linked to the realization of basic infrastructures to get the facility working; other costs could be represented by a piecewise constant function depending on capacity, or by general nonlinear functions of the flows going through that node in the network. Recent trends tend to make some fixed costs variable, as we may lease warehouse space from a provider of logistic services; in a highly uncertain and dynamic setting, this may be an advantage.

Transportation costs present a similar structure, resulting from a mix of fixed and variable costs. When shipping a standard container from a certain point of the supply chain to another one, part of the cost is fixed and independent of the content. Transportation rates may be quite intricate, but again we may find a suitably accurate representation. If we want to compare two transportation modes, we are actually interested in figuring the best solution. If errors in the cost evaluation are not too large, they do not reverse the ranking of alternatives, and we make the correct decision anyway.

More often than not, there is a tradeoff between different cost components. For instance, transportation cost can be reduced by selecting a close supplier; however, this need not lead to the lowest overall cost because, when we order something from a supplier, several factors come into play besides transportation cost:

- ordering costs,
- the price charged by the supplier, which may also be affected by currency exchange rates,
- inventory holding costs.

Unlike transportation costs, ordering costs are *internal* costs, in the sense that they depend on the operations of the buyer firm, whereas transportation cost may depend on either the supplier, or the buyer, or a service provider. In the past, each order was associated with a procedure including some phone calls or fax messages. These costs were largely independent of the amount purchased. This is why we typically consider fixed ordering costs, i.e., associated with the order itself and not with the amount ordered. Electronic commerce has eased this burden considerably, but we may also consider receiving, inspecting, and handling incoming goods as components of the ordering cost. They can be partially captured by a fixed ordering cost. Sometimes, for the sake of simplicity, we aggregate all of the fixed cost components, including transportation, into a fixed ordering cost.

It is not uncommon to compare a geographically close supplier against a distant one who charges a lower price. The decision cannot be taken without specifying an ordering strategy, which is linked to the inventory control policy. The price can also depend on the purchased quantity, as quantity discount opportunities are sometimes offered. Should we take those opportunities? Reducing the purchase cost is certainly attractive, and the possibility of securing a known price might be too, if we fear an adverse movement in prices and/or exchange rates. However, ordering more materials also implies larger inventory holding costs. Inventory holding costs aggregate different cost components. To begin with, whenever we pay for some goods, and these stay in a warehouse for a possibly long time, we have an opportunity cost for the capital tied up in inventories, which we could have invested otherwise. From a financial point of view, too much capital sitting in inventories is bad news. More so, if we had to borrow money to purchase materials. Apart from financial issues, more inventory means more insurance charges, more material handling (with the possibility of wasting materials), larger expenses to heat or to refrigerate the warehouse, etc. If the goods are perishable or subject to obsolescence, we may also face the need of scrapping a significant amount of material: Cisco Systems was reported to take a staggering inventory write-off

($2.5 billion). All of these considerations lead to the idea that inventories should be kept low. Actually, inventory management is all about finding the right tradeoff; we will introduce the well-known Economic Order Quantity formula in section 1.4, to illustrate the tradeoff between cost of inventories and benefits of inventories (i.e., value of the functions performed by inventories).

We close this section by considering costs which may be very hard to quantify, i.e., stockout costs. We have a stockout whenever we run out of stock and we are not able to satisfy demand immediately; this may result in an unsatisfied customer or the stopping of downstream production. In the latter case, the stockout cost may be not too hard to estimate in terms of lost production, but when dealing with customers at a retail store, how much does an angry customer cost? To begin with, the loss of image associated with a stockout is an elusive concept, because it depends on consumer behavior. If we have a stockout and cannot meet an order from a customer, will she wait or go somewhere else? Assuming she is impatient, and the second case occurs, do we lose just this order or the customer altogether? This is very hard to tell; maybe we will never know, because she will just purchase a substitute item without telling anyone. As a further complication, the stockout cost can be linked to the *occurrence* of the stockout itself, or to the *size* of the stockout (e.g., number of customers that could not find the stocked out item). Even if we cannot quantify a stockout cost, we need to keep close control of the service level we offer, trading off other costs against this performance measure. We cover all of these considerations in chapter 5 on inventory management under uncertain demand.

1.3.3 Strategy

After this cursory look at competition factors and cost drivers in supply chain management, it should be clear that there is no way to find a single solution which is optimal from all of the conceivable points of view. In fact, firms adopt quite different strategies. The supply chain for a technologically mature product with a low profit margin must be efficient and inexpensive. In the case of an innovative product, with high margins and maybe a limited life, the overall strategy will be quite different:

- In the retail sector, the availability of goods on the shelves is essential. Still, an unsatisfied costumer might have a negligible impact, particularly for goods which have acceptable substitutes. However, a stockout for a whole product category (say, milk) or for products which are subject to strong brand loyalty (e.g., Coke in the soft drink industry or detergents for personal hygiene) may have serious consequences.

- In the "business-to-business" sector, we may have quite different priorities. Just think of managing the stock of spare parts to replace defective or failed ones in big industrial machining tools. Keeping such machines idle because of lack of spare parts may be extremely costly; indeed, this is a case where stockout costs may be easy to quantify, as there are contracts specifying penalties for lack of service. Quantifying the stockout cost of spare parts for life-critical equipments at hospitals is impossible, but we clearly see that in such a case we need to ensure immediate availability, either by suitable stock levels or, if the cost is too high, by very fast and expensive transportation.

In order to define a strategy, we must associate priorities to competition factors and find cost-effective ways to achieve a given performance target, possibly trading off performance against cost. Firms in different industries will probably define quite different strategies. It is no surprise that managing supply chains for high performance laptop computers requires a different approach than in the case of soap powder. However, even within the same given industry, we may observe quite different strategies.

Example 1.2 Personal computers are sold using different distribution channels, appealing to different consumers. Some consumers are quite sophisticated and want a very specific configuration; they are willing to wait relatively long lead times to get exactly the stuff they want. Others prefer a choice between a few well-defined alternatives, but fast delivery and cheap prices are essential to them. For similar reasons, some consumers do not mind ordering on a web site, whereas other consumers feel much safer buying from more traditional channels, because they want a personal contact in case of trouble with the product. In fact, different market segments can be dealt with by different marketing strategies.[15] □

Example 1.3 IKEA and MC are two dominant players in the Italian retail furniture business. They are both healthy and fast growing companies. However, they have fairly different strategies. IKEA basically designed a self-service environment where customer are asked to select the product they like, take note of the product code, and collect the selected item(s) at the warehouse. IKEA customers tend to transport goods by themselves. IKEA does not provide transportation services (though a business partner located near the counters sells transportation services). IKEA customers are even asked to design their own kitchen through the Internet or at do-it-yourself PC stands in the stores. Moreover, IKEA has a very wide number of product categories ranging from beds and chairs, to carpets and forks. However the range of product designs is rather limited and is dominated by the Swedish minimalistic design. The MC strategy is quite different. Though the prices are comparable, MC only sells furniture. In a MC store one cannot find carpets, forks, etc. However, in a MC store one can find furniture with very diverse designs ranging from classic, to modern, ethnic, romantic, etc. So

[15] See: V.K. Rangan and M. Bell, *Dell Online*, Harvard Business School Case No. 9-598-116, 1999.

the assortment offered by MC is very wide, though in a slightly different way: MC provides fewer product categories, but more styles than IKEA does. Also, while IKEA provides little sales assistance and delivery service, MC has a service intensive strategy. The vast majority of MC customers is attended by a salesperson. A salesperson can spend up to one hour designing the kitchen for a customer that then might simply walk away. Also, 90% of customers ask for the delivery of goods at their place (the cost of delivery is just 7% of the overall price). As we can see, the two companies have very different strategies (in many perspectives they have opposite strategies). However, they both are fairly successful. How can that be? Actually, the key idea is that the two companies appeal to two different segments of consumers and have two different value propositions. IKEA appeals mostly to youngsters (IKEA offers services such as day care for children), who can easily use technologies to design their own kitchen, can transport and assemble furniture on their own, and tend to appreciate the minimalistic Scandinavian style. MC tends to appeal to a more mature population that appreciates more traditional furniture and services such as sales assistance, delivery, and assembly of furniture. □

Perhaps even more surprisingly, the same firm may pursue different operations strategies in space and/or time. In fact, operations may be diversified by geographic region, because alternative markets may require different approaches, depending on customers habits and cultural factors.

Example 1.4 Buying a car follows different patterns on the two sides of the Atlantic Ocean. In the USA, it is common to purchase a car on the spot, after having a look at what is available at the retailer. In Europe, it is more common to order a specific configuration, and possibly wait weeks for the desired model. □

The level of market penetration and/or the potential entry of competitors may also contribute to the definition of a strategy. Finally, time is also essential, as a product at the beginning of its life cycle is typically not managed like an almost obsolete one. For example, a stockout late in the life cycle of a product is almost a desired outcome.

1.4 THE ROLE OF INVENTORIES

Much of what follows in this book deals with inventories; actually, three chapters (4, 5, and 6) are devoted to this topic. Keeping inventories implies a long array of costs, including less obvious ones such as an adverse effect on quality.[16] Indeed, given that many management philosophies are based on the

[16]Quality may be adversely affected because large amount of stocks typically require more material handling, which may result on accidental damage. High inventory levels also

idea of zero inventory, should we could consider inventory management a sort of more or less necessary evil?

Example 1.5 An intuitive consideration is that inventory availability has a positive effect on our ability to satisfy demand. What may be less obvious is that sometimes it is inventory itself that *generates* demand; just think of the allocation of shelf space at a big retail store. Even less obvious, inventory availability may be used to *sense* demand. Consider a large book store. Keeping an inventory of all possible titles is clearly out of the question. However, having some titles covering some discipline may be essential to check if there is potential demand for that kind of book (see case [11]). Otherwise, lack of inventory may imply lack of demand, which may be further motivation for not keeping stock; a perfect vicious circle.[17] Also, some companies keep deliberately large inventories if some staple products to show their dominance and as an implicit promise of product availability, which most customers tend to notice. □

The example above does not imply that we should just increase stock availability. The message is that inventory has a purpose, and that we should understand its role and function in order to plan its level at a facility. The most complex problem is arguably the deployment of stock at the right installation of a large supply chain: on the one hand, we would like to place stock near customers, but this may be the worst place in terms of value of stock, as this is where we have the most added value; furthermore, stock near customers has been committed to a given retail region, potentially reducing flexibility in the allocation of goods. Generally, inventory reduction may be highly beneficial, provided that we eliminate the reasons for keeping it. In order to understand why we might need some inventory, a good starting point is the classical EOQ model.

1.4.1 A classical model: Economic order quantity

In this section we outline a sort of archetypal model for inventory management, the Economic Order Quantity (EOQ) model. Our purpose is just to illustrate how fixed ordering costs affect the need for some stock as well as to lay down some background which will be also used in chapter 2. Hence, the analysis is rather superficial, and much more detail is given in section 4.2.

Consider a distributor selling a good with a rather regular demand pattern. Taking it to the limit, we consider a perfectly constant demand over time. Let

imply longer waiting times on the shelves, which have an impact on perishable items. In manufacturing, high work in process levels are associated with longer flow times; if quality is checked at the end of the process, defects will be detected later, with a possibly significant increase in scrapped material.

[17] For a similar issue, related to phase-in/phase-out of products, see example 3.7 on page 100.

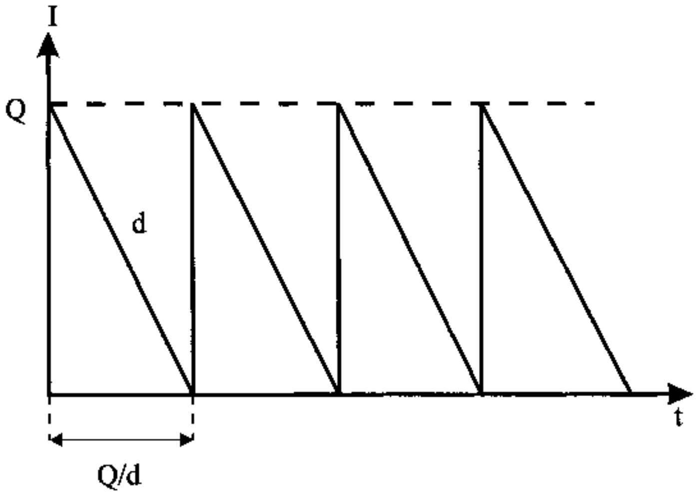

Fig. 1.5 Time evolution of inventory levels in the EOQ model.

d the demand per unit time; the specific time unit is not important, provided that we are consistent in specifying the remaining data (e.g., if demand in measured in units/daym the holding cost is unit of value say euro per unit, per day). The demand must be satisfied from stock, and goods are ordered from a supplier. A natural objective is finding an ordering strategy that allows the distributor to satisfy demand at minimum cost. Given that demand is constant, it is also reasonable to assume that the ordered quantity is always the same. Let Q be the lot size we choose and, without loss of generality, assume that we start with Q units on hand, as shown in figure 1.5. We will run out of stock after Q/d time units. Ideally, we would like to get a new lot of Q parts exactly when the inventory level drops to zero, as this will keep holding cost down. Such a perfect timing is possible if everything is certain and deterministic; this means not only demand, but also the supplier's delivery lead time. If the lead time is denoted by LT, it is easy to see that we should order Q whenever the inventory level[18] drops to a reorder point R given by the demand over the lead time : $R = d \cdot \mathrm{LT}$. If we repeat this cycle over and over, the time evolution of the inventory level will be periodic, as shown in figure 1.5, with cycles repeating every $T = Q/d$ time units.

Let c be the unit price charged by the supplier for each unit; we assume that whenever we order Q units from the supplier, we pay her an amount cQ. In other words, there are no discount opportunities we might take advantage of by ordering a larger amount. We see that cQ is a linearly variable cost. In

[18] In later chapters we will see that ordering decisions should not be simply based on *on-hand* inventory.

the cost c, we could also include a variable component of the transportation cost. Whenever we order, it is also reasonable to expect that a fixed cost has to be paid. This may be due to a fixed component of the transportation cost; or it could be a fixed ordering cost due to the need of issuing and tracking the order. Whatever the case, we denote this fixed ordering cost by A, which does not depend on Q. To summarize, whenever we order Q, the total cost of the order is $A + cQ$. This expression suggests the opportunity of not ordering too often a small amount. We have an economy of scale if we order a larger amount, because the fixed component is distributed on a larger number of parts.

However, there are good reasons to keep Q to a reasonable size. In this very simplified setting we do not consider the risk of obsolescence or perishability, nor physical space limitations in the warehouse. But the least we should do is to consider an inventory holding cost. The simplest reason for dealing with such a cost is the opportunity cost of capital tied up in inventory. There are many other factors which come into play here, but let us simply say that if we keep one part in inventory for a unit period, we face a cost h. Note that the dimensions of this unit inventory holding cost are money per part, per unit time. If we assume that this cost depends linearly on inventory, the total holding cost over some time period is h times the average inventory level.

Example 1.6 We should emphasize that using a linear inventory holding cost, as we will do in most of the book, can be a rather unsatisfactory approximation. To begin with, if we have discount opportunities, we should consider an explicit dependence $h(Q)$; clearly the total opportunity cost does depend on the price we pay per item, and this creates some dependence between the inventory holding cost and the average inventory level, which also depends on Q. Even if we assume that financial costs are more or less linear, other factors may have a nonlinear effect. For instance, consider a very perishable item, whose shelf life is just one day. If we keep inventory levels low, we will probably sell all of the available stock and no material will be scrapped. But if we raise inventory, under demand uncertainty, some leftover inventory will have to be occasionally disposed of. Hence, we see that cost linearity may be a debatable assumption. Still, all of these considerations point out some incentive to keep a low inventory level, maybe within a range such that a linear approximation is acceptable. Anyway, if demand is assumed deterministic, a limited shelf life would simply imply an upper bound on Q, which is easily dealt with. Hence, we will stick to linear holding costs in the following. □

We see that we have two contrasting factors to account for in determining the order quantity Q. To spot the best compromise, we should quantify the total cost per unit time (say, one year) as a function of the decision variable Q. Since inventory level ranges between 0 and Q according to a linear pattern, we see that the average inventory level is $Q/2$. Hence, the holding cost component

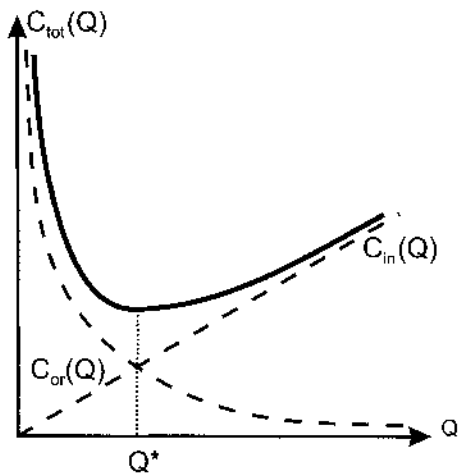

Fig. 1.6 Inventory holding and fixed cost components in the EOQ model.

is

$$C_{\text{in}} = h\frac{Q}{2}.$$

The contribution of the fixed cost component is A times the average number of orders issued per unit time. Since we have d/Q orders per unit time, this component is

$$C_{\text{or}} = \frac{Ad}{Q}.$$

Taking into account the purchasing cost of yearly demand, $C_{\text{pu}} = cd$, we have

$$C_{\text{tot}}(Q) = C_{\text{in}} + C_{\text{or}} + C_{\text{pu}} = h\frac{Q}{2} + \frac{Ad}{Q} + cd.$$

Leaving the last constant component aside, we may draw a qualitative picture of total cost in figure 1.6. We see that the objective function depends on a linearly increasing component C_{in} and a decreasing component C_{or}, displaying an economy of scale with respect to Q. The variable purchase cost plays no role really, as it does not depend on Q under our assumptions (but see example 1.7 below). Now we may find the optimal solution by equating the first-order derivative of the total cost to zero[19]:

$$Q^* = \sqrt{\frac{2Ad}{h}}. \tag{1.1}$$

[19] As we point out in appendix B, this first-order condition need not be sufficient for optimality, and we should also check the second-order derivative or, equivalently, show that the total cost is a convex function of Q. See example B.3 on page 547.

We have just derived the well-known EOQ (*Economic Order Quantity*) formula, which is valid only under a rather long list of limiting assumptions; nevertheless, it provides us with some useful insights. We see that the EOQ size increases with the fixed cost A and decreases with the inventory holding cost h. A simple calculation yields the optimal cost value for the optimal lot size:

$$C_{\text{tot}}(Q^*) = \sqrt{2Ahd} + cd. \tag{1.2}$$

This function shows that the total cost is a *concave* function of demand d; in other words, there is an economy of scale with respect to the demand a facility must face, and we will see in chapter 2 what impact this has on the design of a logistic network.

Example 1.7 (*A remark on relevant vs. irrelevant costs*) Expression (1.2) suggests that, unless discount opportunities are offered, the unit price c we pay for the stocked item is irrelevant in determining the optimal order size. Of course, it is very relevant for the bottom line, since it affects profitability, but we should observe that some costs may be irrelevant when making certain decisions. Actually, a closer look at the formula would suggest that probably c plays some role in determining h. Indeed, a common way to estimate inventory holding cost is to assume some opportunity cost of capital, that is a sort of interest rate i, say 15%, and setting $h = ic$. Nevertheless, the last term cd in (1.2) does look constant and irrelevant in determining Q^*. However, this holds only when we want to select the order quantity Q for a given supplier. If we change the problem at stake, things can change substantially. Suppose that we want to select a supplier and that there are two competitors, whose characteristics are represented by fixed and variable costs, c_1 and A_1, and c_2 and A_2, respectively. When comparing the two suppliers, in terms of the total cost as expressed in equation (1.2), we *cannot* overlook the last cost term. Hence, we see that cost elements and parameters may be irrelevant or not, fixed or not, and this depends on the decision at stake. Indeed, we can tell whether some kind of cost is relevant/irrelevant only with respect to a specific decision. □

Now suppose we wish to reduce the inventory level and/or the corresponding cost. A look at (1.1) suggests that unless we wish to reduce demand or we can reduce inventory holding cost (which increases Q^* but reduces the overall cost), we should reduce the fixed cost A. The fixed cost may depend on the ordering mechanism, the transportation cost, and possibly the setup cost. The setup cost is, within a production context, a fixed charge we pay whenever we start producing an item, independent of how many parts we make.[20] Clearly,

[20] We should note that, in a manufacturing context, setup cost might not be as relevant as the setup time, which reduces machine availability. When capacity is scarce, we cannot overlook interactions among products manufactured using a shared set of resources, and the EOQ model is not well-suited to this task.

if there is such a fixed charge, it is economical to buy or make a suitably large number of parts at once, and this is why inventories may be needed.

This reasoning points out a first function of inventories, which is linked to the need of adapting a relatively continuous and smooth consumption process to a replenishment mechanism, that on the contrary is very lumpy due to purchase, production, or distribution lots. The inventory we build because of this issue is called **cycle stock**. We cannot reduce cycle stock unless we reduce fixed charges, which create the need for a relatively large lots. Indeed, a mainstay of Japanese manufacturing philosophy has been the reduction of setup costs.

However, there are other reasons to build up inventory.

1. Stock is needed to decouple supply and demand, when one of them is subject to variability, and the other one is constrained and cannot follow such variability. In the next subsection we consider how transportation or capacity constraints generate stocks.

2. Stock is needed to hedge against demand uncertainty. The role of demand uncertainty is dealt with later in section 1.5.

We should also mention that there are many more factors that result in the creation of inventories. Raw material stock is sometime created in anticipation of unfavorable market conditions, such as increasing prices or uncertainty in the supply of a scarce commodity. We call this **speculative stock**.

Moreover, it is natural to think of stock as something sitting in a warehouse. However, inventory may be moving, as is the case of **in-transit** or **pipeline** stock. If transportation takes a few hours, in-transit inventory is actually negligible, but if long-distance transportation by ship is used, we may have a non-negligible impact. A similar consideration applies to manufacturing systems: The longer the flow time, the larger the work in process. We should note that while cycle stock depends on the order size, average in-transit stock only depends on average demand and the transportation delay, as illustrated by the following example.

Example 1.8 Consider an Italian firm importing a product from the Far East. The product is transported by ship, which takes one month, and the demand is constant and equal to 1000 pieces per month. If the firm issues a replenishment order once per month, each month it will issue an order for 1000 units, just when the previous one is being received. At each time instant, there is always a ship traveling with 1000 items. If the firm orders once per year, the order size is 12,000 pieces, and during the month following the order (say, January), there will be an in-transit stock of 12,000 items; for the remaining eleven months, in-transit inventory is zero, but its yearly average is still 1,000 anyway. □

Example 1.9 Let us consider a company from the Piedmont region that produces Barolo wine. Let us assume that the company sells 1,000 liters per year. Barolo wine needs to age for at the least three years before it can be sold. Two of these three years need to be spent in oak barrels. Given this demand and these technological constraints, at any point in time the company has at the very least 3,000 liters in stock. To cut this inventory investment either we reduce production volume, or change the technology in order to reduce the three year LT (i.e., find a way to make Barolo wine age more quickly), or simply decide to produce a different kind of wine. □

1.4.2 Capacity-induced stock

In the EOQ model we consider a constant and perfectly predictable demand. However, demand need not be constant to be perfectly predictable. In (very few) lucky cases we may have a time-varying demand which we know, as is the case if we make to order with a long delivery lead time. Ideally, we should be able to deliver all of the items just in time, with no need for stocking end items. As expected, cycle stock might be needed if there are fixed charges in making or buying the items. However, even if there is no fixed charge, we may have to resort to stock items in order to better match demand with capacity.

Example 1.10 Consider an item whose demand is strongly affected by seasonality. For instance, say that average demand is 100 per month, but the actual demand is 200 in spring and summer, and zero in autumn and winter. If items are produced by the firm, rather than purchased from an outside supplier, there are two extreme choices. It may size its manufacturing capacity to the maximum demand (200 units per month). In this case, there is no need for inventory, but capacity utilization is just 50%. At the other extreme, it could size the capacity to 100 items per month. In this case utilization is 100% but there is a considerable inventory buildup during the low-demand season. In this case we speak of **seasonal stock.** □

In figure 1.7 we illustrate a sample time evolution of seasonal stock when capacity is held constant and equal to average demand. It may also be the case that the mismatch is not between constant manufacturing capacity and time-varying demand, but between time-varying raw material availability and constant demand; this is the case for many food goods, such as canned tomatoes and olive oil. Sometimes, one can try to match capacity and demand by producing items with opposed seasonalities. For instance, winter and summer clothing can be produced in the same plants. In other cases, transportation capacity is used as a buffer: Apparently, for half a year kiwi fruit is imported from New Zealand to Italy, and vice versa in the other half.[21]

In a mathematical programming model illustrated on page 542 of appendix B, we illustrate how we might plan inventory buildup in order to match de-

[21] It remains to be seen whether eating kiwi twelve months per year is worth the resulting pollution.

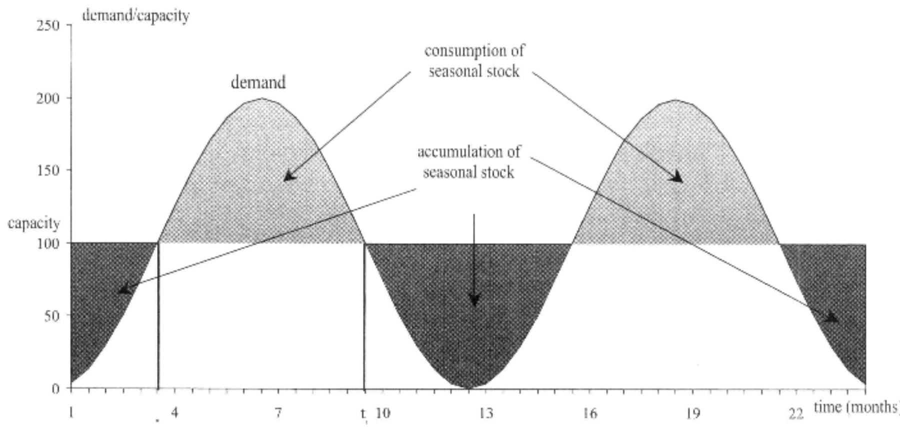

Fig. 1.7 Seasonal stock buildup and depletion.

mand and manufacturing capacity. Clearly, issues may compound with each other: in example B.12 on page 571, we consider a production planning problem where both capacity limitations and fixed charges call for the creation of stock. We should emphasize that besides fixed charges in the form of a fixed cost, we might have fixed consumption of capacity whenever we start production. If it takes a few hours to set up a machine to make an item type, we have to make a fairly large lot not to consume the capacity with setup times. These fairly large lots build up some inventories. This is again a form of cycle stock, even if the motivation is not strictly economical.

In the example above, we have considered production capacity, but in distribution logistics similar considerations apply to transportation capacity. Cycle stock may be necessary if transporting small orders is not economical, but inventory might also be required if the number of vehicles is limited and their capacities need to be fully utilized by full-truckload transportation.

1.5 DEALING WITH UNCERTAINTY

In distribution logistics there are many factors which are significantly affected by some form of uncertainty. For instance, we should extend the EOQ model to account for:

- uncertainty in demand, which may vary according to not perfectly predictable patterns;
- supplier lead time, which is affected both by transportation time and by possible material shortages.

At a different decision level, when tackling a long-period problem, we may have to face uncertainty in:

- prices, both in the sense of prices our suppliers charge and prices we may ask;
- exchange rates, which are relevant in an international context, on both the supply and demand side;
- changes in average demand; e.g., demand might simply fade away because of new emerging technologies.

Uncertain factors may be different in nature, depending on the length of the time horizon on which decisions must be made. Furthermore, different types of uncertainty may compound; for instance, demand uncertainty may be the result of short-term random variations in demand level, or of more systematic factors such as the success of a product and its market penetration, which also depends on the behavior of competitors.

In fact, we may consider different concepts of uncertainty. The probabilistic concept of uncertainty, which may be modeled by random variables following a given probability distribution, is the most common one. Other paradigms have been proposed, but we will essentially stick to a more familiar statistical framework (see chapter 5). If we know the relevant probability distribution, then we just have uncertainty in the realization of random variables. More often than not, properties of random variables must be inferred from available data, assuming they are available and reliable. In such a case, we have some uncertainty as far as the probability distribution itself, or its parameters, are concerned. Nevertheless, if data are available, we are still in the domain of probability and statistics and deal with a sort of "objective" uncertainty. In extreme cases, we deal with a brand new innovative product, and past information is simply unavailable, or its relevance might be questioned. In that context, we have to deal with subjective assessments of uncertainty (see section 3.12).

Whatever the nature of uncertainty, we must come up with some way to mitigate its effects. In the next two sections we consider two examples illustrating the role of safety stocks and proper product design.

1.5.1 Setting safety stocks

We have illustrated *how much* we should order according to the EOQ model, but we should also clarify *when* we should order (for a more detailed discussion see section 5.4). If both demand and supplier lead time are constant, it is easy to see that we should order an amount Q whenever the inventory level falls below a reorder point $R = d \cdot LT$ corresponding to the demand during lead time. In doing so, we should consider not only the physical (on hand) inventory, but also orders that we have already sent but have not been delivered yet, and backorders. Under deterministic assumptions, items will be delivered exactly when on-hand inventory reaches the zero level. When uncertainty is involved in either demand or lead time, or both, it is intuitive that we should raise the reorder level (in most situations, for a more detailed discussion see chapter 5). To do so rationally, we need two ingredients:

1. a description of the uncertainty of demand during lead time;
2. a suitable definition of the quality of service we want to offer our customers, in terms of our ability to meet demand immediately from stock.[22]

The uncertainty of demand during lead time depends on how the two basic uncertainties, in demand per unit time and in the lead time itself, are compounded. A typical assumption is that it can be modeled by a random variable D_{LT}, with normal distribution, expected value μ_{LT}, and standard deviation σ_{LT}.[23]

As far as the service quality is concerned, we will see in chapter 5 that different measures can be reasonably defined; we could also set up an optimization model, provided we may quantify the cost of a stockout. We consider here the simplest, not necessarily the best, alternative, which is to set a constraint on the probability of a stockout. This probability, denoted by α, should be suitably small; correspondingly, we define the quantity $1 - \alpha$ as our service level. Typical values of the service level could range between 90% and 99%. We have a stockout during lead time if demand in that time span exceeds the reorder point R. The probability of not having a stockout is

$$\text{P}\{D_{\text{LT}} \le R\} = 1 - \alpha.$$

We immediately see that R is the $1-\alpha$ quantile of a normal distribution with parameters μ_{LT} and σ_{LT}.[24] As shown in appendix A, calculating the quantile of an arbitrary normal distribution boils down to finding the corresponding quantile for a standard normal distribution. Knowing the quantile $z_{1-\alpha}$ for a standard normal variable, we set

$$R = \mu_{\text{LT}} + z_{1-\alpha}\sigma_{\text{LT}}.$$

The idea is illustrated in figure 1.8: The shaded area, on the right of the quantile, corresponds to the stockout probability α. In the deterministic case, we simply set $R = \mu_{\text{LT}}$; doing so when lead time demand is normally distributed would result in a 50% service level. In order to increase the service level, we add a **safety stock** given by $z_{1-\alpha}\sigma_{\text{LT}}$. Clearly, safety stock increases the overall cost. On the average, we have an additional inventory of $z_{1-\alpha}\sigma_{\text{LT}}$

[22] In a make-to-stock system or retail environment, this is the fundamental ability. In a make-to-order, quoting a reliable lead time may be more relevant, whereas in assembly-to-order, the customer should be allowed to customize her order in an easy and flexible way.

[23] Central limit theorem may justify such an assumption in the case of consumer goods; see page 470. This hypothesis should be checked by a suitable statistical procedure (see section A.9.1).

[24] See definition A.9 on page 456.

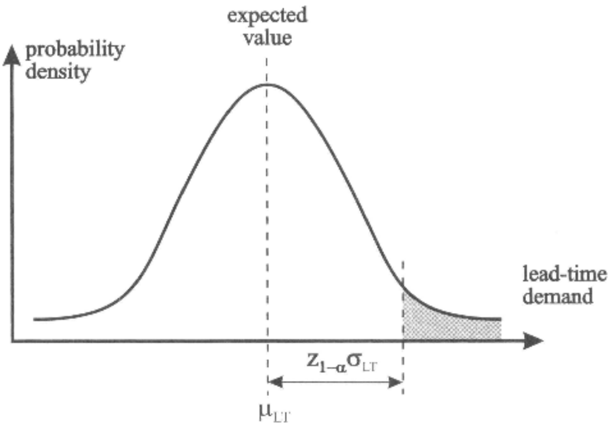

Fig. 1.8 Calculating safety stock based on lead-time demand uncertainty.

parts, with respect to the EOQ model; hence, if we set the order quantity according to the EOQ model[25] and we disregard variable purchasing cost, the average total cost (1.2) becomes

$$\sqrt{2Ahd} + hz_{1-\alpha}\sigma_{LT}. \tag{1.3}$$

We should also mention that in this expression we are not considering stockout costs, which will be essential in chapter 5. Looking at equation (1.3), we see two sources of cost: cycle and safety stock. Setting a safety stock is, in some sense, a *passive* answer to the problem of uncertainty; we simply add some slack resources to reduce the effects of demand uncertainty. We could try to be proactive and prepare a set of actions to reduce the need for a large safety stock. Reducing safety stock, without reducing uncertainty is just not a good option, unless we want to give up service quality or ignore stockout costs. This calls for reducing uncertainty in lead time demand. On the one hand, lead time should be reduced; in a deterministic setting, the lead time might be irrelevant, because in that case we have just to anticipate the order timing (yet, it could be relevant in terms of in-transit inventory). In an uncertain setting, while the average lead time LT does not enter explicitly equation (1.3), it contributes to increasing σ_{LT}. We will see in chapter 5 that, if demands during different time periods are independent random variables, then σ_{LT} increases with the square root of lead time: $\sigma_{LT} = \sigma\sqrt{LT}$, where σ is the standard deviation of the demand per unit time. We will also see, in chapter 6, how demand and lead time uncertainty can be compounded.

[25] We will see later, in chapter 5, that this need not be the optimal choice. We should select the two parameters Q and R jointly, and perhaps consider an alternative definition of service level, taking into account the size of the stockout, and not only its occurrence.

It is tempting to believe that demand uncertainty is out of our control, and there is nothing we can do about it. Sometimes, this is true, but in many cases demand uncertainty is not exogenous. In many cases we can simply reduce demand variability that creates uncertainty. Demand spikes can be the result of unanticipated promotional sales; indeed, some large retail chains have decided to avoid promotional sales altogether, adopting an *every-day low prices* (EDLP) policy. In other cases, we may try to improve our forecasting procedures in order to (partially) transform unpredictable variability into predictable variability. This is basically the purpose of forecasting techniques described in 3.

In complex systems, many other actions can be attempted to reduce safety stock costs. In manufacturing systems, we may adopt preventive maintenance policies in order to reduce the occurrence of random machine breakdowns, and the related need for inventory buffers. In a multiechelon system, we may try to hold inventories in central warehouses where demand is more aggregate and thus more predictable, rather than at retail stores. In the next section, we illustrate the general idea of *risk pooling* by exploiting common components in assembling end items.

1.5.2 A two-stage decision process: Production planning in an assemble-to-order environment

When computing safety stocks, we do not plan orders in advance: We prescribe the structure of a policy (e.g., Q and R), and we let the system run and place orders when our policy suggests to do so (e.g., when we hit the reorder point R). In practice, the parameters are adjusted periodically. Furthermore, emergency actions are carried out when needed. All of these adjustments are carried out when additional information is obtained, but this is outside this formal model. The formal model is, in a sense, single-stage: We make some decisions and then see what happens. In some other cases, we want to include in a formal model the adjustments we might make at a later stage. In order to do so, we must formalize the dynamics of the decision process, whereby decisions are made and/or revised when new information is obtained. This may lead to very difficult stochastic models. We consider here a simple example of a two-stage model.

Consider an assemble-to-order (ATO) system. In such a system, we have to make (or buy) components, which are then assembled into some end item we sell. It would be nice to do everything after we receive a customer order, but we cannot afford this luxury if the customer is not willing to wait that much time. If the customer wants everything immediately, we have to keep a stock of end items; this may be difficult or impossible when end items come in a wide variety of configurations or when items cannot be stocked because of their cost. A compromise solution is feasible when making components requires a long lead time, but assembly is relatively fast. We can keep a stock of components, which are made or bought *before* we get customer orders. We assemble only on order, i.e., after we collect customer demand. Concrete examples of ATO processes are the automotive industry, at least in Europe, and the PC industry, where one can order a customized model and select among a number of feature/options.[26]

Let us build a simple but instructive model along with a small numerical example, under the following assumptions:

1. First, we decide how many units of each component we build, subject to manufacturing capacity constraints. This first decision sets the total production cost.

2. After receiving customer orders, we use components to assemble finished goods. The assembly plan in designed to maximize revenues; the cost term in the profit function is fixed by the previous decision (if we neglect assembly cost); if components are not enough to meet customer orders, we lose profit opportunities; if too many components are available, they are discarded with a possibly considerable loss of money.

The key point, apart from demand uncertainty, is that we have a limited time window for sales, after which components are no use. This is a limit assumption, typical of the classical newsvendor model (see section 5.2): in practice, components might have some salvage value, or they could be used in later time periods. In this setting, we have to make two decisions in sequence, in order to optimize profit. Literally, we cannot maximize profit, because it is a random variable depending on our decisions and on uncertain demand, but we may maximize its expected value.[27]

Since the main complicating factor is demand uncertainty, one possibility is to disregard it and just use expected values of demand in planning production of components. Another possibility is representing demand uncertainty by a set of scenarios. We will pursue both approaches and compare the decisions we make.

To set up a small toy example, say that we own a (very) small firm, producing just three end items (A_1, A_2, A_3), which are obtained by assembling components (c_1, c_2, c_3, c_4, c_5). The components we use for each end item are described by a bill of materials, which is flat (just two levels: end items and components). The bill of materials is given in the left-hand side of table 1.1. From the bill of materials, we see that there are two common components, c_1 and c_2, while the remaining three are specific and characterize each end item. We assume that three resources (M_1, M_2, M_3) are used for production of components. On the right-hand side of the table we also see the bill of resources,

[26] A possibly more pleasing example is any pizzeria offering a wide array of pizzas; the pizza is made on order, but all of the components are prepared in advance.

[27] A more sophisticated approach would involve some considerations about risk, which is not fully captured by the expected value.

Table 1.1 Bill of materials for the assemble-to-order example

	c_1	c_2	c_3	c_4	c_5
A_1	1	1	1	0	0
A_2	1	1	0	1	0
A_3	1	1	0	0	1

Table 1.2 Bill of resources, cost of components, and available capacity

	M_1	M_2	M_3	Cost
c_1	1	2	1	20
c_2	1	2	2	30
c_3	2	2	0	10
c_4	1	2	0	10
c_5	3	2	0	10
Cap.	800	700	600	

Table 1.3 Demand scenarios, expected value of demand, and selling price of end items

	S_1	S_2	S_3	Exp. Demand	Selling Price
A_1	100	50	120	90	80
A_2	50	25	60	45	70
A_3	100	110	60	90	90

i.e., the time required on each resource to manufacture one component. In the table, we also give the available capacity for each resource type, and the cost of each component; this cost might include both direct variable production costs and material costs. We assume that assembly is not a bottleneck, hence its capacity is disregarded.

Other relevant data concern end items, demand, and the price at which end items are sold. They are given in table 1.3. Demand uncertainty is modeled by a set of three scenarios (S_1, S_2, S_3). If we have information about past sales, the three scenarios may result from the discretization of a continuous probability distribution (of course, more scenarios are needed in a practical setting to approximate the distribution); alternatively, they could result from an interview with three experts. Whatever the case, we assume that the three scenarios are equally likely, i.e., each probability is 1/3.[28] We also give the expected value of demand, which is obtained by averaging the three scenarios for each end item. The last column displays the price at which end items are sold.[29] Also, note that the selling price is larger than 60, the total component

[28] When discretizing continuous distributions, we might use different probabilities to get a better approximation; see. e.g., [4, chapter 10] for an application of Gaussian quadrature. In the case of forecasts based on subjective judgment by experts, using the same probabilities means that we consider three equally reliable experts.

[29] If we do not want to disregard assembly cost, we may substitute selling price by contribution to profit from assembling and selling an item; this defines the second-stage cost, as it takes selling price and assembly cost into account, but not component costs.

cost, for all of the three end items, but A_3 looks more profitable, in a sense, because its profit margin including component costs is $90 - 60 = 30$, whereas A_2 is the least profitable; of course, this reasoning may be misleading in that it does not take into account resource consumption.[30]

We may tackle the problem of maximizing expected profit by the Linear Programming (LP) techniques described in appendix B. To build a simple model as a starting point, we could disregard uncertainty and deal with one scenario characterized by average demand. We get the following model:

$$\max \quad -\sum_{i=1}^{5} C_i x_i + \sum_{j=1}^{3} P_j y_j, \tag{1.4}$$

$$\text{s.t.} \quad \sum_{i=1}^{5} T_{im} x_i \le L_m, \qquad m = 1, 2, 3, \tag{1.5}$$

$$y_j \le \bar{d}_j, \qquad j = 1, 2, 3, \tag{1.6}$$

$$\sum_{j=1}^{3} G_{ij} y_j \le x_i, \qquad i = 1, 2, 3, 4, 5, \tag{1.7}$$

$$y_j, x_i \ge 0.$$

Here, subscript i refers to components, subscript j refers to end items, and subscript m refers to resource types. Input data correspond to those reported in the tables:

- the component cost C_i;
- the selling price P_j for each end item;
- the available capacity L_m for each resource type (measured in time units);
- the resource requirement (processing time) T_{im}, for component i on resource m;
- the number G_{ij} of components i going into an end item j (i.e., the bill of materials - BOM);
- the expected demand $\bar{d}_j$, which is assumed certain.

The decision variables are x_i, the number of components of type i that we produce, and y_j, the number of end items of type j that are assembled and sold; to be more precise, we *pretend* that we will really sell an amount y_j, because we disregard demand uncertainty. The model aims at maximizing profit, as expressed by the objective function (1.4), subject to capacity constraints (1.5). The inequality (1.6) states that we cannot sell more than what

[30]See example B.1 on page 537.

is demanded, whereas (1.7) says that we cannot assemble end items if the required components are not available. The decision variables are required to be non-negative. In fact, for the sake of simplicity, we consider a *continuous* LP model, which allows for fractional quantities; if we insist on requiring that produced and assembled quantities are integer, it is easy to incorporate this requirement (see section B.6).

Solving the model, e.g., by the simplex method (see appendix B), we get the following solution (rounded to two decimal digits):

$$\begin{array}{lll} x_1^* = 116.67, & x_2^* = 116.67, & \\ x_3^* = 26.67, & x_4^* = 0.00, & x_5^* = 90.00, \\ y_1^* = 26.67, & y_2^* = 0.00, & y_3^* = 90.00. \end{array}$$

In this very small example, we may easily interpret what this solution tries to accomplish. We assemble the maximum number of end items of type A_3, which is the most profitable one; this requires in turn the production of a corresponding number of common components c_1 and c_2, as well as the specific component c_5. Since demand limit is binding for A_3, there is some capacity left, which is used to produce a limited amount of the specific component c_3, which is needed to assemble end item A_1, plus common components. End item A_2 has the lowest selling price and is disregarded, as well as is its specific component c_4. It should be noted that, in general, one should not take for granted that the production of the highest profit item should be maximized; the consumption of available resources should be taken into account as well (see example B.1 on page 537 for a counterexample).

In this specific case, the solution is quite readable, but it is a bit "extreme." An expert planner would immediately see that it is a risky bet on high sales of the most profitable item. The optimal profit, according to this model, is 3233.33, but this is actually misleading. After planning production of components, we do *not* know the value of profit, but only its distribution (if we accept the validity of the demand scenarios). We cannot maximize optimal profit; what we can do is maximizing its *expected value*, and this requires a more sophisticated model that takes demand scenarios into account:

$$\max \quad -\sum_{i=1}^{5} C_i x_i + \sum_{s=1}^{3} \pi^s \left(\sum_{j=1}^{3} P_j y_j^s \right), \tag{1.8}$$

$$\text{s.t.} \quad \sum_{i=1}^{5} T_{im} x_i \le L_m, \qquad m = 1,2,3, \tag{1.9}$$

$$y_j^s \le d_j^s, \qquad j = 1,2,3, \quad s = 1,2,3, \tag{1.10}$$

$$\sum_{j=1}^{3} G_{ij} y_j^s \le x_i \qquad i = 1,2,3,4,5, \quad s = 1,2,3, \tag{1.11}$$

$$y_j^s, x_i \ge 0.$$

The big change in this model, with respect to the expected demand model [(1.4)–(1.7)], is that demand uncertainty is taken into account explicitly. Here we consider demand d_j^s for item j in scenario s. Accordingly, the quantity assembled is now represented by scenario-dependent decision variables y_j^s; this is the amount of end item we assemble and sell, if and when scenario s is realized. Assembly decisions are not taken here and now, when we plan production of components, but they are *contingent plans*. The scenario-independent variables x_i are first-stage variables, whereas variables y_j^s are second-stage variables. So now we implement the production plan (i.e., first stage decisions x_i) and develop a contingency plan for the assembly operations (i.e., second stage decisions y_j^s). Only when demand is realized we choose among the contingency plans (y_j^1, y_j^2, y_j^3).[31] We should carefully notice the difference between a multiperiod model and a multistage model. We illustrate examples of multiperiod models in appendix B and in chapter 4. In such models, decisions will be implemented in later time periods, but they are all taken *now*, based on the currently available information. It is possible to revise such decisions by solving the model again according to a rolling horizon strategy, but this is outside the scope of the model itself. In a multistage model, we do not commit to one specific decision for the later stages; the decision that will actually be implemented depends on the realization of random variables, and it will be fixed only when the relevant information will be available in the future. Next-stage variables may also be used to "adjust" previous decisions, given current contingencies. This interpretation explains why models such as the one above are called stochastic programming models with recourse.

Going into details of the model above, the objective function (1.8) consists of a first-stage (deterministic) term accounting for the cost of components, along with a second-stage term, which is the expected revenue from selling end items (not including component cost); the expected value is computed by summing the revenues under the three possible decisions, times scenario probabilities π^s. The capacity constraint (1.9) is unchanged, because it pertains to first-stage only. The market constraint (1.10) is now scenario-dependent, as it considers the stochastic demand d_j^s. Finally, constraint (1.11) links the two stages, stating that assembly is constrained by component availability, for each end item and each scenario. Solving this model, we get the following solution:

$$x_1^* = 115.71, \qquad x_2^* = 115.71,$$

[31]Notice that this holds only when the three scenarios are actually the only three possible demand scenarios. In other cases, we can face a very large number of different scenarios (possibly an infinite number of different scenarios). In this case, the three scenarios are only meant to model demand uncertainty and make sure that first stage decisions account for demand uncertainty. The realized demand might differ from all three scenarios. In this case, once demand is realized we simply have to write a second model for assembly decisions, where we need to meet the realized demand with a limited quantity of components that was fixed through the above model.

$$x_3^* = 52.86, \qquad x_4^* = 2.86, \qquad x_5^* = 62.86,$$
$$y_1^{1*} = 52.86, \qquad y_2^{1*} = 0.00, \qquad y_3^{1*} = 62.86,$$
$$y_1^{2*} = 50.00, \qquad y_2^{2*} = 2.86, \qquad y_3^{2*} = 62.86,$$
$$y_1^{3*} = 52.86, \qquad y_2^{3*} = 2.86, \qquad y_3^{3*} = 60.00.$$

The real outcome of the model is the set of the first-stage decision variables x_i^*. Observing the component production plan, we immediately see a qualitative difference with respect to the model disregarding uncertainty: It is less extreme. We do not produce a large amount of component c_5, because we do not place a risky bet on high sales of A_3. In fact, scenario three would prove a disaster for the deterministic solution: In that scenario, sales are lower for A_3, but we could not react because we do not have enough specific components for the other end items. This also implies that many specific components[32] would be thrown away (according to our assumptions concerning the limited time window for sales and the lack of any salvage value of unused components). The stochastic model, instead, increases production of specific component c_3, which is needed to support assembly and sales of A_1; even a small amount of component c_4 is produced, in order to support the least profitable end item A_2, which helps in using common components when sales are low for other end items. While there is a big difference in terms of specific components, we see that as far as common components are concerned, the solutions of the deterministic and the stochastic solutions are essentially the same. There is a good reason for this, as common components are a flexible resource, which can be exploited to support different end items. Moreover, the demand for common components is the sum of the individual demands for the end items, and by aggregating demand we often reduce uncertainty. Indeed, this *risk pooling* effect is what we try to exploit in assemble-to-order systems. In chapter 6 we will see that the same mechanism is exploited in the management of distribution networks. However, it is also important to note that when end item demands are strongly correlated, the risk pooling effect is considerably reduced. In such a case, we should expect that even the produced quantities of common components differ in the deterministic and the stochastic model. Another relevant factor is capacity: If this is so tight that we may sell whatever we are able to produce, a simple deterministic model could be a viable option.

But how do the two solutions compare in terms of profit? The objective function from the solution of the second model is 2885.71; apparently, the stochastic solution is worse than the deterministic solution, whose objective value was 3233.33. But this comparison makes no sense. We are actually comparing two different situations rather than two different solutions. The above finding simply proves that we would rather face a certain demand rather than an uncertain one. The objective function of the first model is neither

[32] In the more general case even common components could be thrown away.

the true profit, which is uncertain, nor its expected value. It would be the optimal profit, if we knew that the average demand scenario is what will be realized. In the first model [(1.4)–(1.7)] we *pretend* to know the end item demand, and we get the illusion of higher profits. In order to compare the two solutions, we should fix the production plans for components suggested by the two models, and then we should solve a set of second-stage problems, where we optimize assembly of end items subject to component availability, for different demand scenarios. More formally, given a set of first-stage variables x_i^* for components, we should solve the following second-stage (recourse) problem for each scenario s:

$$
\begin{aligned}
R^s(\mathbf{x}^*) \equiv \quad \max \quad & \sum_{j=1}^{3} P_j y_j^s, \\
& y_j^s \le d_j^s, \qquad j = 1, 2, 3, \\
& \sum_{j=1}^{3} G_{ij} y_j^s \le x_i^*, \qquad i = 1, 2, 3, 4, 5, \\
& y_j^s \ge 0,
\end{aligned}
$$

where $R^s(\mathbf{x}^*)$ is the optimal revenue we collect under scenario s, given the first-stage solution $\mathbf{x}^*$, and making optimal use of available components to meet demand. The first-stage solution can come from solving a stochastic or an expected-value model; whatever the case, its expected revenue is

$$\sum_s \pi^s R^s(\mathbf{x}^*).$$

Expected profit for an arbitrary solution can be obtained by subtracting its first-stage cost from its second-stage expected revenue.[33] To evaluate the deterministic solution, we should plug it in this model; in case of scenario S_1, the optimal assembly and sales plan is

$$y_1^* = 26.67, \qquad y_2^* = 0.00, \qquad y_3^* = 90.00,$$

and the same holds for S_2. The bad news is that if scenario S_3 occurs, we are in trouble, because the high-risk solution does not fit demand very well. The optimal assembly and sales plan would be

$$y_1^* = 26.67, \qquad y_2^* = 0.00, \qquad y_3^* = 60.00.$$

This is a pretty bad scenario with low sales and corresponding low profit. We must compute revenue for each scenario, multiply it by its probability,

[33] We are evaluating expected profit *in-sample*, i.e., by using the same set of scenarios which are used in the stochastic model; we could use a much larger set of out-of-sample scenarios to get a more reliable estimate. The point is that solving a large number of small LP problems may take less CPU time than solving one large stochastic LP model.

sum everything to get the expected value, and subtract the component cost from the first stage. Doing so, we may see that the expected profit from the deterministic solution (2333.33) is much lower than what the objective function of the deterministic model [(1.4)–(1.7)] predicts (3233.33), based on one average-case scenario. The percentage improvement of the stochastic solution with respect to the deterministic one is

$$\frac{2885.71 - 2333.33}{2333.33} \approx 23.67\%.$$

Clearly, we cannot extrapolate general results from a small toy example. Indeed, the advantage of using a stochastic model is striking here, because specific components have a large impact. In a case featuring a lot more component commonality, the result would be less impressive. Furthermore, we have assumed that unused components are scrapped, which need not be the case. They could have some salvage value, and we could have a multistage problem so that they can be used in later stages. Nevertheless, the example is quite instructive in pointing out:

- the difference between decision *stages* and *time periods*,

- the role of risk pooling.

In this case, risk pooling is obtained by using common components and by deferring assembly decisions. To further illustrate the value of deferring decisions in a more specific distribution setting, some fashion retail chains send only a part of the items to retail stores at the beginning of a season; at a later stage, after observing sales at each retail store, the residual stock is sent downstream. Also in this case, the first decision, i.e., the purchase of items from suppliers, is often constrained by a budget assigned to each *buyer* in charge of a specific market segment. The second decision, inventory allocation, can be made by a different type of professional called *planner*.

A last important consideration, which applies to all models we describe in this book to deal with demand uncertainty, is that we have considered the maximization of expected profit as a suitable objective. We do not consider profit variability across scenarios, or what happens in extremely bad scenarios (the average smoothes out single outcomes). This makes sense if we may repeat the game over and over (for various items or over multiple periods), so that what really matters is the average profit in the long run. However, in the short run we may take too many chances: If a single bad decision cannot be recovered, because we immediately go out of business, or get fired, a more careful approach should be taken to fully account for risks. An alternative view, for economically minded readers, is that considering expected profit is equivalent to assuming a risk-neutral attitude; risk-averse decision makers should consider different objective functions.

1.5.3 Inventory deployment

The previous section serves well to illustrate the role of commonality in order to reduce the impact of uncertainty. Common components mitigate uncertainty by providing flexibility and by allowing postponement of critical decisions. This is just one instance of the more general risk pooling concepts which are widely used in distribution logistics. When we consider an arborescent network like the one in figure 1.3, we should decide if and how much safety stock we should place at each node. We will see in section 2.1.1 that placing safety stocks upstream may reduce their aggregate level.[34] On the other hand, we should be careful to ensure suitable customer service, which would require locating stock downstream. We see that the inventory deployment decision is by no way trivial, and as usual there is no ready answer for all possible circumstances. As the following example shows, creative thinking may be required in peculiar cases.

Example 1.11 Consider the problem faced by a manufacturer of very expensive spare parts for some industrial equipment[35]; the manufacturer itself, or a firm providing maintenance services in its place, signs a contract requiring immediate replacement of faulty parts, say within a few hours. Where should spare parts be located, and how many of them are required? The second question requires possibly nontrivial probabilistic modeling. As far as the first question is concerned, allocating one part to each customer would certainly ensure satisfactory customer service, but it would be extremely costly. One alternative could be to place some stock at a facility which is more or less located in a barycentric position with respect to customers. However, if a customer is far, we should probably arrange for very fast transportation, maybe by air. With very fast transportation, the exact location of stock may be irrelevant. Hence, we could even consider placing spare parts at some customer location, reserving the right to collect the part for fast shipment to another customer in need of a spare part. This would save some warehouse cost, but it requires a shift in the paradigm prescribing that the owner of stock is the owner of the location where the inventory is placed. The spare part changes owner only when it is mounted on a machine. □

The example illustrates a simple case of a more general strategy called Vendor Managed Inventory, which is later illustrated in example 1.12 on page 41. For reasons that will be later explained in chapters 6 and 7, it may be advantageous to have only one authority in charge of inventory management, since the interactions of different decision makers having limited information, and typically misaligned incentives, may generate unwanted spikes in demand; this phenomenon is known as the bullwhip (or Forrester) effect. In fact, it

[34] See section 2.1.1.
[35] See [8], page 611.

is important to keep in mind that managing a complex supply chain is not just a technical challenge, as human factors and different points of view may exacerbate difficulties (see chapter 7).

The possibility of postponing inventory allocation decisions and exploiting risk pooling depends on product design too. The supply chain of HP DeskJet printers was successfully reorganized by changing the assembly process,[36] in such a way to delay differentiation of products (e.g., according to destination country). For example, they assemble the printer with instruction manuals, cables, and plugs at the warehouse, rather than at the production site. This may result in an increase in the direct product cost, but the analysis must be carried out on a global level, taking into account the shorter and shorter life cycle of products, whose obsolescence may be very fast (indeed, this is the case in consumers' electronics). While this assembly process might add a few cents to the direct production cost, customizing products in the central warehouse might cut inventory investment and obsolescence cost by millions of euros. Generally, demand forecasting is easier whenever we may aggregate items by family. Consider clothing, which may differ in model, size, and color. If we may postpone dyeing items, in order to gain more reliable information about demand, considerable savings may be obtained. Indeed, a well-known case in this vein is Benetton, where cutting and dying operations were swapped in order to ease forecasting.[37]

1.6 PHYSICAL FLOWS AND TRANSPORTATION

In section 1.2 we have considered a network as a physical arrangement of facilities. An essential feature of any supply chain is the selection of a transportation strategy and the management of physical flows, inbound and outbound from any node. Large organizations manage transportation by themselves, whereas in other cases this activity is outsourced; in general, we should decide between alternatives such as rail, ship, air, or trucks.

Restricting our attention to road transportation, we may arrange point-to-point transportation or route a vehicle to serve multiple destinations. For instance, referring to figure 1.3, we may have one vehicle for each link from node 3 to nodes 6,7, and 8; alternatively, the same vehicle may visit the three retail stores sequentially. A decision problem that may occur in the first case is the determination of a suitable transportation frequency; in section 2.1.2 we show that a simplified version of the problem, accounting for fixed and inventory holding costs, closely resembles the EOQ model. In the second case, we should find a suitable assignment of customers to vehicles, and a

[36] This case study is described, e.g., in [10].

[37] See S. Signorelli, J.L. Heskett. Benetton (A). Harvard University Business School Case, 1984.

customer sequence for each vehicle, in order to optimize a given performance measure; such a problem, known as the Vehicle Routing Problem, is dealt with in chapter 8.

When operating our own vehicles, we may try to utilize their capacity at best, according to a *full truckload* strategy (e.g., see the case [7]). Sometimes, the need for fast delivery requires *less-than-truckload* (LTL) transportation. For instance, fast mail couriers typically cannot easily exploit full transportation capacity (trucks and aircrafts), and they try to aggregate flows by proper design of the transportation network. In the LTL case, we may also consider the use of third-party transportation, leaving to our business partner the task of aggregating flows in order to better exploit capacity.

1.7 INFORMATION FLOWS AND DECISION RIGHTS

In figures 1.1, 1.2, and 1.3 we have illustrated the flow of goods, but the information flow is just as important. In principle, information pertaining to the whole supply chain can be collected and managed by a unique decision maker. This centralized manager, should be able to come up with globally optimal decisions. Information Technology (IT) might make all of this a concrete possibility, but there may be unsurmountable difficulties. To begin with, an all-encompassing decision model may be way too difficult to solve. A nastier difficulty is the reliability of information. All large retail stores use point-of-sale data acquisition, and we should be able to know exactly how much stock is available and where, for each item. In practice, such information need not be 100% reliable because of errors, theft, wrong deliveries on the part of suppliers, misplaced inventory, exceeded shelf-life, damage due to material handling, etc.

Even leaving the above difficulties aside, there are deeper difficulties with a fully centralized decision-making architecture:

- Actors in the supply chain may be unwilling to share information.
- Actors in the supply chain may be unwilling to relinquish decision rights to others.

Example 1.12 The **Vendor Managed Inventory** (VMI) approach is a good case to illustrate difficulties in information sharing and allocation of decision rights. Consider a supplier, who delivers goods to independently owned retail stores. Point of Sales (POS) information can easily be collected and sent to the supplier, who could plan inventory accordingly. By the same token, retailers should send timely information to the supplier in case of planned promotions; otherwise, unpredicted demand spikes may have both immediate consequences, such as stockouts, and long term ones, such as loss of customers to competitors, which further contribute to the difficulty in forecasting demand an planning inventory. In fact, a retailer who receives a reduced amount of stock, because of a shortage, may be tempted to order more than needed during the next replenishment cycles, in anticipation of rationing strategies on the part of the supplier. But if all of the demanded items are eventually delivered, a low-demand period will follow because the retailer must get rid of excessive stock. This contributes to an increase in demand volatility along the supply chain, as well as to an overall feeling of partner unreliability. These and other reasons contribute to the generation of the so-called **bullwhip effect**, which has been well-known since the 1960s (see section 6.3). One way to overcome this issue would be to centralize demand information from POSs, which can be collected by the supplier. While technically possible, this solution may be thwarted by retailers feeling that the supplier could share this information with their competitors. An even more radical approach is based on the idea that the supplier is not only the collector of all information in the supply chain, but also the only actor in charge of managing stocks. In VMI, goods are stocked at retail stores, but they are managed by the supplier and change owner only when goods are placed on the shelves. A very well-known case in this vein is Barilla,[38] a firm that had to work very hard to persuade retailers to adopt such a policy and give up authority on inventory. □

A general issue raised by VMI is: Assuming that a (maybe partially) centralized policy reduces the overall costs, who is going to enjoy the benefit? More generally, if multiple actors (different firms, or separate branches within the same firm) control different managerial levers along the supply chain, is there any guarantee that the overall strategy is optimal? There is no easy and general answer to such very delicate issues. In chapter 7 we clarify the related issues and outline the design of incentives to improve overall performance. Given the complexity of the involved issues, that chapter has more of a conceptual than operational nature.

1.8 TIME HORIZONS AND HIERARCHICAL LEVELS

In distribution logistics we have to tackle quite different problems in terms of time horizon, involved uncertainty, and impact of the decisions we make. Designing a new network of warehouse facilities, to be operated during the next few years, and organizing vehicle routes for the delivery of the next day are clearly two extreme examples of problems pertaining to different hierarchical levels.

- At the highest hierarchical we have **strategic** problems. The time horizon may be years or months. The longer the time horizon, the higher the level of uncertainty, which calls for suitable forecasting procedures

[38] See J.H. Hammond. Barilla SpA (A). Harvard University Business School Case, 1994. Alternatively, the case is described in [13].

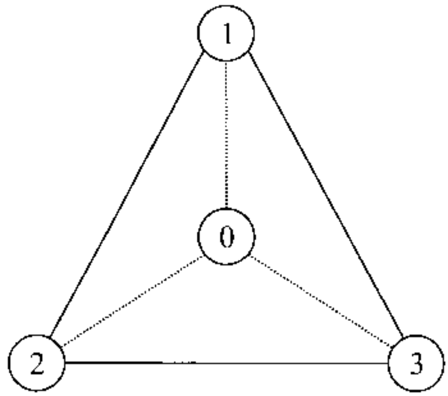

Fig. 1.9 Graphical illustration of the location problem in example 1.13.

and scenario analysis. Decisions made at the strategic level, such as warehouse capacity, will play the role of constraints at lower levels in the hierarchy.

- At an intermediate level we have **tactical** problems. Here, resource availability is usually fixed, but the time horizon (say, weeks) is long enough to require some form of forecasting. An example of tactical problem is the selection of an inventory management policy; changing such a decision is definitely easier than redesigning the structure of a distribution network.

- At an **operational** level, we have day-to-day decisions, where uncertainty is negligible, and we have to react to incoming information in a very short time span.

It is worth noting that the division between the three levels is not sharp at all. Third-party providers of logistic services allow us to enlarge warehouse space without building any new facility, and this makes the boundary between strategic and tactical problems less clear.

Furthermore, the hierarchical levels are interdependent. Of course higher-level decisions constrain lower-level management, but the link is two-way. Hierarchical decomposition is needed to tackle otherwise intractable problems, but when making a strategic decision we must somehow anticipate the effects on the tactical and operational levels. This must be done by some model simplification, resulting in a sort of "anticipation" function. We will see a few examples in the next chapter, but the next example illustrates how different decisions cannot always be taken disregarding their interactions.

Example 1.13 Consider a network consisting of three retail stores, located on the vertices of an equilateral triangle, as illustrated in figure 1.9. The position of nodes 1, 2, and 3 is given, and we should locate a distribution node (or a production plant) in such a way that the total transportation cost from the distribution to the retail stores is minimized. Assuming that the demand on the three retail stores is the same, an intuitive solution would be placing the distribution center in the barycenter of the triangle (node 0 in the figure; the resulting vehicle routes are drawn as dotted lines). However, this depends on the transportation mode. If demand is large with respect to vehicle capacities and we transport point-to-point, this solution is reasonable. However, if demand is low and distances are not too large, it could be much better to visit nodes 1, 2, and 3 in sequence with the same vehicle. In such a case, we could place the required node at the same location of any retail point on the perimeter and, in case it saves us some money, we can consider to place it in one of the three vertexes of the triangle (stores). Of course, a real-life problem should also account for fixed cost components in transportation, environmental issues, item perishability, etc. □

We close this section by stressing again the fundamental difference between *time periods* and *decision stages* (see section 1.5.2). In a long-term decisions, we may prepare plans which are implemented in successive time periods. This may result in dynamic problems. If the decisions are made here-and-now and are not changed later, we to have a multiperiod decision problem, but it is a single-stage one. We typically reserve the term "multistage" for problems in which future decisions are adapted as a function of additional information we gather and of the progressive resolution of uncertainty.

1.9 DECISION APPROACHES

Supply chain management strategies may differ according to priorities in objectives, information availability, and strategy of the firm (see section 1.3). There is wide array of possibilities, and confusion is sometimes added by ambiguous use of buzzwords, such as **push** vs. **pull** systems. Indeed, the manufacturing literature has largely contributed to this state of the matter, because of the confusion among different hierarchical levels, such as demand management (also known as master production schedule in a production environment) and shop-floor control.

The following classification criterion is suggested in [13, chapter 5]:

> *In a push-based supply chain ... production decisions are based on long-term forecasts. (...) In a pull-based supply chain ... production is demand driven .... rather than..forecast (driven).*

We may also substitute "purchasing" or "distribution" for "production," to make the definition more general. So the difference between a push and a pull strategy is the following:

- In a pull system, purchasing, production, or distribution orders are based on the consumption of a good in the downstream operation. For example, in a manufacturing environment, the production of a component might be triggered by the consumption of that component at an assembly plant. In a distribution environment, the distribution of a case-pack of canned tomatoes is triggered by the consumption of canned tomatoes at the stores. These policies are somehow based on minimum inventory levels and once they are reached the upstream stages start purchasing, producing or distributing the products.

- In a push system, purchasing, production, or distribution are based on a plan, which is based on a forecast of a future demand. For example, in a production environment we might decide to produce a component because our assembly plan for next week foresees the need for such a part. Also, in a distribution environment we might distribute a large quantity of a given product, because we foresee a peak in demand due to a promotion.

To be fair we shall say that somehow pull strategies are based on some sort of forecast as well. Indeed, while materials consumption triggers purchase, production, or distribution orders, these are governed by parameters that are based on some sort of forecast.

Example 1.14 Now, let us consider a simple EOQ-based policy, where a replenishment order is issued when we reach the reorder point. Is this a pull policy? From a certain point of view, it certainly is: We issue an order when inventory is pulled. However, some forecasting procedure is arguably used in setting the policy parameters, which depend on the expected value of demand and its standard deviation. We see that even a simple pull approach is based on a mix of demand forecast and materials consumption. The parameters are based on some sort of forecast, while the replenishment orders are triggered by actual demand and materials' consumption. □

Example 1.15 Kanban production control can give us another good example of a pull system. Kanbans are a means to control production at the shop floor level; they were invented in Japan and made famous by Toyota. Kanbans are basically a permission to start the production. These "permissions" to produce are released only when the components are actually consumed. So they are a very effective means to control the inventory level. If we only have permissions for 100 units (say we have 100 kanbans and each gives the permission to produce one unit), we never have more than 100 units of the component at stake. When a unit of the component is consumed in the assembly operation, we release the permission to produce one unit. On the other hand the manufacturing stage attaches one kanban to each unit manufactured. Therefore, while the consumption of the component releases "permissions" to produce, production consumes them. This process makes sure that the inventories of components do not get out of control, since the manufacturing stage can produce one unit if and only if one unit has been consumed. This makes the kanban production control the gold-standard for pull systems. However, one could wonder why we decided to have 100 kanbans rather than just 50? On the other hand, one might wonder whether 100 kanbans are enough? Actually, these decisions may be made by simple rules of thumb, simulation models, or even complex algorithms that lie outside the scope of this book that focuses on logistics. However, one can intuitively understand that the number of kanbans depends on the rate of consumption of the components, which in turn depends on the expected future demand for the finished good. Another relevant factor is uncertainty, which provides us with an incentive to raise the number of kanbans in order to add some safety stock. □

Quite often, "pull" is associated with a good and efficient policy, whereas "push" is associated with obsolete practice. Actually, in some contexts the pull strategy might perform very poorly whereas the push strategy might be very effective.

Also, there is nothing like a pure strategy, as real-life approaches are typically hybrid mixtures, and these terms should be associated to *features* of a solution approach, rather than to a specific choice. So the key issue is not choosing between one strategy and the other. We rather have to find the right blend at the various levels, as the two examples below show.

Example 1.16 For example, in many supply chains, we develop long-medium term plans to allocate resources, plan shifts, give suppliers advance notice of expected changes in demand, etc. For example, a company might sign a contract for the supply of 10,000 cans of Coke a month, based on the expected demand over the next 3 months. Nevertheless, the actual delivery-orders might be driven by the actual consumption of Coke at the stores; for example, stores might reorder a pallet of Coke when the inventory level of Coke drops below a given threshold. So we have a long term push strategy, whereby we commit to the overall quantity based on some sort of forecast of future demand. On the other hand, the short term replenishment process is driven by the consumption of Coke at the stores and thus can be considered to be pull. As we can see, push and pull are features of the solution rather than contrasting alternatives. □

Example 1.17 In a production environment we might have a master production schedule (that is the plan for production of the finished good) where we plan the production quantities over time, according to current inventory levels, future demand (either a demand forecast or firm orders or a mix of the two), setup costs, etc. This is actually a plan at the finished good level. So one would be led to think that companies that use the Master Production Schedule use a push strategy. Actually, at the shop floor level the replenishment of components to be assembled might be driven by their consumption and thus might fall under the "pull umbrella." The assembly of finished products is based on a schedule, whereas (some) components are produced and replenished as they are consumed by assembly operations. This example too shows that the push and the pull logics can coexist and very often are used by the same company. □

A second recurring theme in operations is the Make-to-Stock/Make-to-Order dilemma. First we should realize that **Make-to-Order** is not a synonym of pull system and **Make-to-Stock** is not a synonym of push system. An example will, hopefully, make the point clear.

Example 1.18 Let us go back to the car industry (see example 1.4). Both in the USA and in Europe the replenishment and production of components is based on a pull strategy, at the least in the short run as the kanban production control has become a sort of standard in this industry. Nevertheless, in the USA most cars are made to stock, while in Europe they are made to order. This clearly shows that push or pull can be associated with either Make-to-Stock or Make-to-Order. □

When one thinks carefully about it, the issue is actually fairly simple. The flow of components to the assembly line can be based on a pull or a push strategy, but the fact is that both strategies simply disregard whether a specific customer (say Mr. Brandimarte) is waiting for the blue car, with leather seats, and air conditioning, or the car is simply ordered by a retailer (or a commercial unit) that hopes to sell it sooner or later to a generic consumer.

Also, Make-to-Stock and Make-to-Order are not actually contrasting alternatives, but they should rather be considered as features of a strategy, and can be combined to design a reasonable solution. For example, in most good restaurants dishes are prepared to order, while raw materials are purchased to stock. Things are fairly easy for standard raw materials with a long shelf life such as flour or potatoes. Things are more tricky for very specific and short shelf life products such as mullets (a specific kind of fish that is used for very specific recipes). They are bought if and when we expect that on the same day (or the next day) a customer will ask for a very specific recipe.

Moreover, the assemble-to-order example of section 1.5.2 suggests the possibility of integrating different strategies. Components may be produced (or purchased) based on forecasts, whereas final assembly is made only when a customer order is received. As we pointed out, this is a necessary arrangement when the delivery lead time accepted by the customer is smaller than the overall lead time for producing the end item, but it is impossible to stock end items, because of their cost or their variety. We see that there is an *order decoupling point*[39] which separates subsystems governed by different policies. Finally, it is also worth noting that quite different approaches may be adopted within the same firm, depending on specific items (in terms of value, perishability, etc.) and customers.

[39] Sometimes, the term "push-pull boundary" is used in manufacturing.

Example 1.19 Consider for example a manufacturer of top-end watches. The basic models (maybe still worth a few thousand euros) are available at the stores. On the contrary, unique items such as top grand-complication items (that is items with an extremely complex mechanical movement that can account for lunar phases etc.) are made to customer order as demand is so sparse that it makes no sense to carry them over. Also these extremely expensive items are only bought by collectors that seem to enjoy the time they have to wait, as it testifies the product is really hard to make and is specifically made for them. □

1.10 QUANTITATIVE MODELS AND METHODS

In this book we use quantitative models and methods extensively. Applying a quantitative approach means setting up a mathematical model and solving it by some appropriate method. The quantitative feature could be associated to some "scientific" or "objective" virtue, but this is a somewhat reductionist approach. As the saying goes, there is no such a thing like an exact model: All models are wrong, but some are useful. This is why modeling has been defined as the art of selectively simplifying reality. Choosing the right degree of simplification is indeed an art, which is subject to often contrasting views depending on personal taste and opinion. Since building and solving a model is done with some purpose in mind, different stakeholders may have quite different ideas about the right modeling approach. Whatever the case, there are many reasons making simplification necessary:

- Computational tractability: As we point out in appendix B, some optimization models may be hard to solve, and we must give up some modeling detail and/or resort to suboptimal solution methods.

- Uncertainty: In principle, we may use the machinery of probability theory and statistics (see appendix A) to represent uncertainty, but sometimes lack of data, or difficulty in the model, prevents an exact representation. We should also keep in mind that not all of the uncertainties can be formalized within the framework of probability theory.

- Complex dynamics may prevent elegant analytical modeling.

- There are often conflicting points of view, which cannot be analyzed objectively on a purely quantitative basis.

There are two wide classes of quantitative models:

1. Prescriptive models. Typical examples are optimization models, which are formulated with the aim of getting a decision directly. In principle, decision-making could be automated by gathering data, instantiating a mathematical programming model, and solving it by one of the many

commercial solvers. In practice, prescriptive models should just be used as a decision support tool.

2. Descriptive models. Unlike prescriptive models, modeling tools within this class do not aim at generating a decision. They just try to capture relationships between variables, shedding some light on key features of the problem at hand, which is then used by the decision maker.

While we will illustrate many quantitative models, we should emphasize that useful descriptive models may also be qualitative; their role is rationalizing a business process and reaching a common understanding, which is not to be taken for granted in large organizations or in contexts involving several firms with different views and incentives.

The descriptive models we consider in this book are mainly aimed at predicting something. Prominent examples that we will consider are time-series-based forecasting and regression models (see chapter 3). We might also consider performance evaluation models. The idea is predicting the performance of a real system, for a certain configuration and for a given setting of some parameters governing decision rules. To make this point a bit more concrete, let us denote by $f(\boldsymbol{\theta}; \omega)$ a performance measure depending on a set of decision variables $\boldsymbol{\theta}$, which are under our control, and a set of random variables, which are beyond our control; the dependence on random events is expressed by ω. A performance evaluation model aims at estimating the expected value of the selected performance measure:

$$H(\boldsymbol{\theta}) \equiv \mathrm{E}_{\omega}[f(\boldsymbol{\theta}; \omega)].$$

Performance evaluation models may further split into two subclasses:

1. Analytical models.

2. Simulation models.

Analytical models typically require some simplification. We will see some examples in chapter 5 when deriving approximations of expected cost as a function of inventory management policies under uncertainty. Analytical models in this domain may require some simplifying assumptions; for instance, we will assume that backordering is possible, i.e., customers are patient. But if customers are not necessarily willing to wait, demand can be lost, making modeling harder. Simulation models, on the contrary, are extremely flexible and powerful, at least in principle; however, they require much effort in data gathering, and maybe in solution time, and require a working knowledge of general-purpose programming languages or more specific simulation environments. While in other engineering-related problems we need continuous-time simulation models, in supply chain modeling we need *discrete-event* simulation models. By "discrete-event" we mean that the system state changes in correspondence with specific events: For instance, the inventory level changes abruptly when a supplier delivers an order, or when a customer asks for some material. Uncertainties are modeled by pseudo-random number generators, i.e., algorithms able to emulate random phenomena, such as customer demand. The simulation program includes event management and decision rules which allow us to emulate the time evolution of quantities of interest and to estimate required performance measures given the set of parameters $\boldsymbol{\theta}$. There are graphical description languages, which may make the modeling task easy in simple cases, as they require assembling and linking standard blocks with a graphical editor; still, nontrivial thinking may be required to fit a complex system within the bounds of the selected simulation environment.

Because of these reasons, we do not deal with simulation modeling in the book, but we want to point out that, as usual, we should not draw a very thick line separating prescriptive and descriptive models. For instance, many revenue management and dynamic pricing strategies use regression modeling (e.g., to capture the link between price and demand), as well as modern optimization software tools. Furthermore, modern simulation environments are integrated with optimization solvers able to manage simulation experiments in order to automatically search for the best setting of parameters with respect to a specified cost or profit function. Since there is randomness in any supply chain, we actually want to optimize (say, maximize) the *expected* value of some performance measure:

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} H(\boldsymbol{\theta}) \equiv \mathrm{E}_{\omega}[f(\boldsymbol{\theta};\omega)],$$

where $\boldsymbol{\Theta}$ is the feasible set for the controlled parameters $\boldsymbol{\theta}$. The expected value is, when random variables are continuously distributed, a possibly high-dimensional integral. Then, we must resort to some sampling mechanism, yielding an approximation $\hat{H}(\boldsymbol{\theta}) \approx \mathrm{E}_{\omega}[f(\boldsymbol{\theta};\omega)]$. For simple systems, we may get an analytical approximation, which is suitable for optimization by mathematical programming, as we have seen in the stochastic optimization example of section 1.5.2. When simulation is needed, we have to resort to different optimization approaches. Typically, commercial software relies on some form of evolutionary computing able to deal both with noisy estimates of the performance measure and with usually nonconvex optimization problems.

1.11 FOR FURTHER READING

- In this book we will deal with problems which lie at the boundary between distribution logistics and production planning. An excellent book on manufacturing systems, including production planning and control, is [8].

- An excellent text covering supply chain management with a wider scope (and, necessarily, sometimes a more shallow level) is [5]. Among other things, the reader will find there some treatment of revenue management and electronic commerce. For a text very rich in references to practical cases, see also [13].

- We deal with distribution logistics from an *operations management* perspective, but we should keep in mind that this dimension must be linked to a financial perspective; models integrating the two sides of the coin are illustrated in [12].

- We have pointed out that there is no best supply chain management approach; the strategy must be adapted to the specific firm and market at hand, a point which is very well illustrated in [6].

- Readers interested in discrete-event simulation will find [9] very comprehensive and readable.

- A tutorial introduction to stochastic programming models in manufacturing can be found in [1]. For a comprehensive introduction to both models and solution methods, see, e.g., [3].

REFERENCES

1. A. Alfieri and P. Brandimarte. Stochastic Programming Models for Manufacturing Applications. In A. Matta and Q. Semeraro, editors, *Design of Advanced Manufacturing Systems.* Springer, Dordrecht, 2005.

2. C. Billington, G. Callioni, B. Crane, J.D. Ruark, J.U. Rapp, T. White, and S.P. Willems. Accelerating the Profitability of Hewlett-Packard's Supply Chains. *Interfaces*, 34:59–72, 2004.

3. J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming.* Springer-Verlag, New York, 1997.

4. P. Brandimarte. *Numerical Methods in Finance and Economics: A MATLAB-Based Introduction (2nd Ed.).* Wiley, New York, 2006.

5. S. Chopra and P. Meindl. *Supply Chain Management: Strategy, Planning, and Operation (2nd Ed.).* Pearson Prentice Hall, Upper Saddle River, NJ, 2004.

6. M.L. Fisher. What Is the Right Supply Chain for your Products? *Harvard Business Review*, 75:105–116, 1997.

7. P. Ghemawat and J.L. Nueno. *Zara: Fast Fashion, case 9-703-497.* Harvard Business School Publishing, Boston, MA, 2003.

8. W. Hopp and M. Spearman. *Factory Physics (2nd Ed.)*. McGraw-Hill, New York, 2000.

9. A.M. Law and D.W. Kelton. *Simulation Modeling and Analysis (3rd Ed.)*. McGraw-Hill, New York, 1999.

10. H.L. Lee and C. Billington. Material Management in Decentralized Supply Chains. *Operations Research*, 41:835–847, 1993.

11. A. Raman and Z. Ton. *Borders Group Inc., case 9-601-037*. Harvard Business School Publishing, Boston, MA, 2003.

12. J.F. Shapiro. *Modeling the Supply Chain*. Duxbury/Thomson Learning, Pacific Grove, CA, 2001.

13. D. Simchi-Levi, P. Kaminsky, and E. Simchi-Levi. *Designing and Managing the Suppy Chain (2nd Ed.)*. McGraw-Hill/Irwin, New York, 2002.