Service Inventory Management

solution techniques for inventory systems without backorders

Bijvank, Marco, 1981, -
Service Inventory Management: solution techniques for inventory systems without backorders
ISBN 978-90-8659-386-6

Thomas Stieltjes Institute
for Mathematics

Cover photo: grocery store in Seattle, United States by Marco Bijvank

Printed by: PrintPartners Ipskamp, The Netherlands

VRIJE UNIVERSITEIT

Service Inventory Management

solution techniques for inventory systems without backorders

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op donderdag 22 oktober 2009 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Marco Bijvank

geboren te Goes

promotor: prof.dr. G.M. Koole
copromotor: dr. I.F.A. Vis

To my proud father.

Preface

Performing research and writing this thesis has been an exciting and very fulfilling job over the last couple of years. Many people believe that you have to be really smart to write a thesis. Normally, you hear Ph.D. students complain about the amount of work, the courses they have to take and teach, the lack of guidance, the long office hours, etc. All this is true. However, there is also another side to performing research. There are periods of time during which you are really busy. But you are also able to spend your time on something you love to do most, and set up your own time schedule when to do it. In the mean time you participate in conferences all over the world to present your work. This has resulted in trips to Brussels, Paris, Brescia, Budapest, Reykjavik, Seattle and Washington. I have also been to Aarhus for two months to perform research. These experiences have offered me the possibility to meet new and interesting people. I would like to take this opportunity to thank all people who have supported me through these years and made my experiences much more enjoyable.

The first persons to mention are my supervisors. Ger gave me the opportunity to start my Ph.D. trajectory. Not only did I enjoy my time at the VU University, I will never forget the hiking trips in the Alps under his guidance. Iris helped me a lot on how to perform research and she taught me the writing skills that I have today. I am very grateful to have had both of them as my supervisors. During my visit to Aarhus University, I have learned a lot from Søren Johansen. In the final stage of my thesis, the reading committee (consisting of Sandjai Bhulai, Geert-Jan van Houtum, Tim Huh, Ruud Teunter, and Steef van der Velde) has read my thesis carefully. I am thankful for their suggestions to improve my thesis.

I also would like to show my gratitude to all of my colleagues, and several in particular. Sandjai Bhulai has been a true friend and neighbor in the office. We have had some fruitful discussions on my research. We shared great conversations and many laughs. Being a Ph.D. student has also been a great experience due to my roommates. Wessel van Wieringen has persevered me for about three years, and the other way around! Furthermore, I would like to thank my fellow Ph.D. students. Vivek, Wemke and myself kept each other motivated to finish our theses.

On a closer note, I am thankful to my family and friends for their interest and support. It has been a rough path with many ups and downs. I am sure that it has not always been easy for Sonja to live together with me. During some periods she had to cope with unusual “office” hours. I will always love her for her patience and understanding, her empathy to distress me on a personal level, her

bright spirit and the woman she has been to support me at all times.

Marco Bijvank
September 2009

Contents

Chapter 1

Introduction

Advertisements are found everywhere to persuade customers into buying a certain product. Due to the globalization of the market the number of brands for similar products is growing and the amount of competition is huge. For example, as a customer you can go to a local bookstore to buy a book. But on the internet customers can find more titles for lower prices. In the latter case, customers even have the convenience of staying at home. Therefore, retailers have to differentiate themselves from their competitors. Customer service is one of the most important differentiation strategies for retailers (see, e.g., Daugherty [68] and the references therein). Customers are usually satisfied when they get what they expect at the moment and in the quantity they want it.

All kinds of parties are involved in fulfilling a customer's request. The system of organizations, people, activities, information and resources involved in the production, transportation and sale of a particular product is called a *supply chain*. Supply chain activities transform natural resources, raw materials and components into a finished product that is delivered to the end customer (Nagurney [215]). It includes the manufacturers, suppliers, distributors, wholesalers, retailers and end customers themselves as depicted in Figure 1.1. In a supply chain the resources and information can go from supplier to customer and the other way around. This is called downstream and upstream, respectively. The primary purpose of most supply chains is to satisfy customer needs while generating profits for all parties involved in the process.

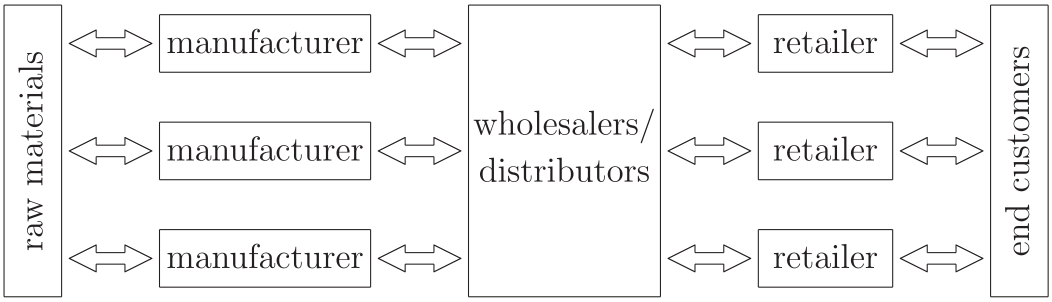


Figure 1.1: A supply chain including the product and information flow among all parties.

Recently, there is a shift in supply chains from a focus primarily on manufacturing and costs to a focus that increasingly includes a services management element (see, e.g., Boyaci [43]). In the former strategy customer service is implicitly included in terms of costs, whereas in the latter strategy customer service is explicitly included in terms of a target service level. Truly focusing on end customer needs requires excellent forecasting or an extremely responsive supply chain. Inventories are kept at all parties of the supply chain to incorporate flexibility against uncertainties. Traditionally inventory management tends to look for cost-based improvements, rather than focusing on customer service. This shift towards customer focus should also be reflected in inventory management. The goal of this thesis is to develop inventory models and to design related solution approaches which resolve around customer service. We call this *service inventory management*. Current trends from practice that have to be dealt with are tight delivery schedules, efficiency and cost reductions, service constraints, increased competition and the influence it has on customer behavior. Current solution techniques do not address these issues sufficiently (as will be indicated below). Therefore, new approaches and techniques are required to control inventory levels more accurately. This is the goal of this thesis. Basic inventory theory concepts and standardized terminology are introduced in Section 1.1. The scope and contribution of this thesis are discussed in Section 1.2. An outline of this thesis is given in Section 1.3.

1.1 Introduction to inventory control

Inventory is the number of products or resources held available in stock by an organization and can include raw materials, work-in-process, component parts, and finished products. The inventory of manufacturers, distributors, and wholesalers is clustered in warehouses. Retailers keep their inventory either in a warehouse or in a store accessible to customers. Many types of inventory exist:

- *safety stock* is the amount of inventory kept on hand to protect against uncertainties in customers' demand and supply of items. The reason to keep this type of inventory is because demand and lead times are not always known in advance and have to be predicted. This type of inventory is also called *buffer stock*.

- *seasonal stock* is the inventory built up to anticipate on expected peaks in sales or supply, such that the production rate can be stabilized. This is also called *anticipation stock*.

- *cycle stock* consists of the inventory waiting to be produced or transported in batches instead of one unit at a time. Reasons for batch replenishments include economies of scale and quantity discounts.

- *decoupling stock* is used to decouple the output of two inter-dependent workstations because of different processing rates, set-up times or machine breakdowns. This permits the separation of decision making.

- *congestion stock* results from items that share the same production equipment. Consequently items have to wait for workstations to become available and inventory is built up.

- *pipeline stock* includes inventory in transit between different parties of the supply chain. This is also called *work-in-progress*.

Not every type of inventory is kept at all parties in the supply chain. For example, decoupling stock and congestion stock are mainly kept in a manufacturing environment, whereas safety stock and pipeline stock are more important in a retail environment. In order to handle the different types of inventory, control systems have to be developed. According to Hax and Candea [110], an *inventory control system* is a coordinated set of rules and procedures that allows for routine decisions on *when* and *how much* to order of each item in order to meet customer demand. A *replenishment policy* specifies how to decide upon these two decision variables. A classification of inventory control systems is given in Figure 1.2 and will be explained below.

Most inventory systems concern *single-item* systems and consider one type of product at a time. In *multi-item* inventory systems a number of products are considered simultaneously because of limited capacity availability, economies of scale for joint replenishments or other reasons. The classification for single-item systems is described in the remainder of this section. A similar classification can be made for multi-item systems and is therefore not included in the figure. See for more details Zipkin [327]. Furthermore, we talk about an *inventory model* when it represents the inventory system. A model can include assumptions and, therefore, it is a simplification of reality.

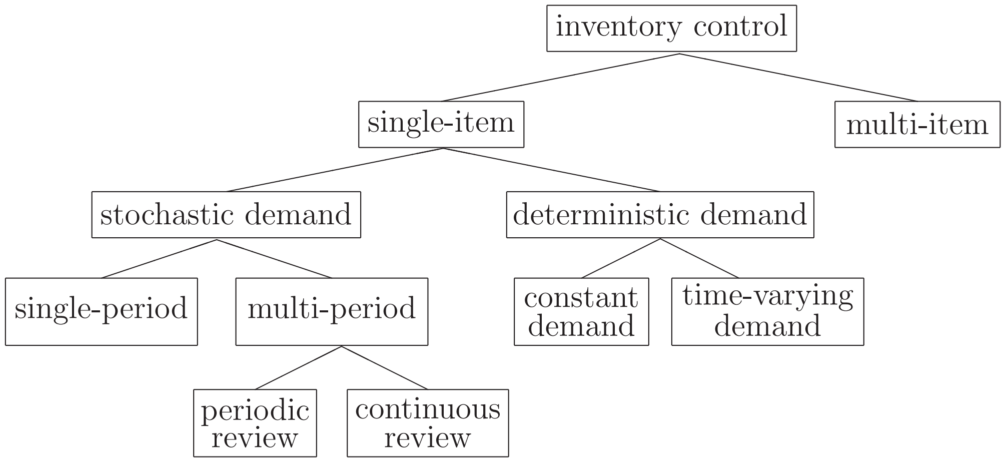


Figure 1.2: A classification of inventory control systems.

There is a distinction between predictable and unpredictable demand. If the demand in future periods can be forecasted with considerable precision, it is reasonable to use an inventory policy that assumes that all forecasts will always be accurate. This is the case in *deterministic* inventory models. However, when demand cannot be predicted very well, it becomes necessary to use a *stochastic* inventory model where the demand in any time period is a random variable rather than a known constant. Models in which demand is known (or forecasted) and constant over a planning horizon are called *classical lot size models*. A well-known example is the economic order quantity (EOQ) model introduced by Harris [107]. When demand is not constant but still predictable, the models are called *dynamic lot size models*. Examples of techniques to determine the order size for such models are the Wagner-Within algorithm [303] and the Silver-Meal heuristic [267].

In real life demand is mostly not known in advance. Therefore, a probability distribution is used to describe the behavior of the demand in stochastic inventory models. A special class of inventory control systems is concerned with products which have a very limited period before it can no longer be sold. Examples are perishables (like food and flowers) or items with a limited useful life (like newspapers and fashion). For such inventory systems no decision has to be made regarding the order moment when the replenishment should take place, but only the order size has to be determined for a single time period. Such models are called *single-period* inventory models or *newsboy models*. In *multi-period* inventory models it should also be determined when replenishment orders are triggered.

Different inventory levels are considered to determine when an order has to be placed. The *on-hand inventory* level is the amount of physical inventory immediately available on the shelves in a store or warehouse to meet customer demand. The occasion when the inventory level drops to zero is called a *stock out*. The demand exceeding the available stock is referred to as *excess demand*. There are two ways to deal with this demand when there is a stock out. First, if the customer is willing to wait, the excess demand is held until the next delivery replenishes the inventory. This is called *backlogging* or *backordering*. Alternatively, the customer may not be willing to wait. In this case excess demand is lost, which is called *lost sales*. The on-hand inventory minus the backorders is called the *net inventory*. A positive net inventory represents the inventory on hand whereas a negative net inventory refers to a backlog. The *inventory on order* is the work-in-progress or the items ordered but not yet delivered due to the lead time. When there are backorders, a part of the inventory on order is already reserved to meet customer demands from the past. Therefore, the *inventory position* is defined as the sum of the inventory on hand plus the inventory on order minus the outstanding backorders (or backlog).

How often the inventory status should be checked for replenishments is specified by the *review interval*. This is the period that elapses between two consecutive times at which the stock level is known. Two types of review systems are widely used in business and industry. Either inventory is continuously monitored (*continuous reviews*) or inventory is reviewed at regular periodic intervals (*periodic reviews*). The former type of control is often called *transaction reporting*, since continuous surveillance is not required but only at each transaction that changes the inventory position (e.g., demand or order delivery). Whether or not to order at a review time is determined by the *reorder level*. This is the inventory position at which a vendor is triggered to place a replenishment order in order to maintain an adequate supply of items to accommodate current and new customers. The mathematical notation for the reorder level equals s. It comprises the safety stock and the quantity of stock required to meet the average demand during the lead time plus the time until the next review moment. The *lead time* is the period of time between order placement and the delivery of the order such that the order is available for satisfying customer demands. An order size can either be fixed or variable. The type of replenishment policies with variable order quantities are called *order-up-to policies* in which the order size is such that the inventory position is increased to an order-up-to level. This level is denoted by S. Figure 1.3 shows the difference between continuous and periodic reviews when the order size is a fixed number and each customer demand equals one unit (also called *unit-sized demand*). In the continuous review case, the order is placed immediately when the inventory position reaches the reorder level. In the periodic review case, the order placement has to wait for the next review time after the inventory position has reached the reorder level. Figure 1.4 illustrates the concept of order-up-to policies in the case where customer demands are not always unit sized. Notice that the delay in the actual order placement can result in larger order sizes in case of periodic reviews as compared to continuous reviews.

The mathematical notation for the four discussed types of replenishment policies is shown in Table 1.1. The letter R specifies the length of the review

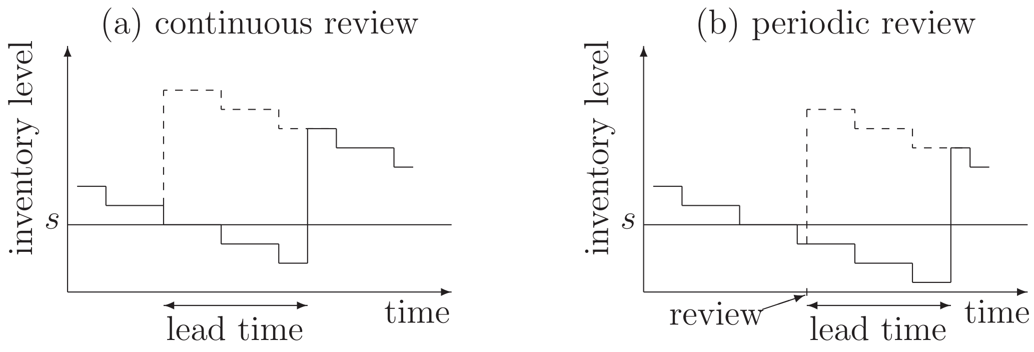


Figure 1.3: The on-hand inventory level (solid line) and inventory position (dashed line) for replenishment policies with a fixed order size under (a) continuous and (b) periodic review.

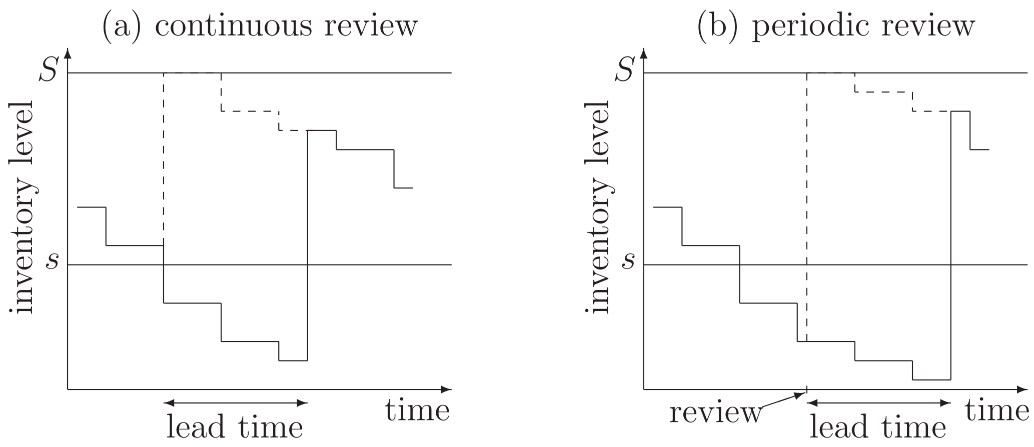


Figure 1.4: The on-hand inventory level (solid line) and inventory position (dashed line) for order-up-to replenishment policies under (a) continuous and (b) periodic review.

interval. The review interval length for continuous review systems is zero and is therefore omitted. Furthermore, s is the reorder level, Q stands for fixed order quantities, and S denotes the order-up-to level. Figure 1.3a and Figure 1.3b illustrate the (s, Q) and (R, s, Q) policy, respectively, whereas Figure 1.4a and Figure 1.4b give an example of the (s, S) and (R, s, S) policy, respectively. A special class within the order-up-to policies are *base-stock policies*, in which the satisfied demand in between two review times is immediately ordered at the next review time. For such policies, the reorder level s is equal to the order-up-to level minus one in case of discrete demand. When demand is continuous, the reorder level is equal to the order-up to level. Periodic review models with a base-stock policy are denoted as (R, S) policies. In continuous review models, base-stock policies are also called *one-for-one policies* since every customer demand immediately triggers a new order. Such models are denoted as $(S-1, S)$ policies. Notice that the (s, Q) and (s, S) policy are identical if all demand transactions are unit sized and $S = s + Q$. In that case replenishments are always made when

		order moment	
		continuous review	periodic review
order size	fixed	(s, Q)	(R, s, Q)
	variable	(s, S)	(R, s, S)

Table 1.1: The notation for the types of replenishment policies most often applied in literature and practice.

the inventory position is exactly at the reorder level. Consequently, the order size always equals Q or $S - s$. See Figure 1.3a for an illustration.

To compare the performance of the different replenishment policies, the costs associated with each controlling system has to be minimized while simultaneously meeting a desired customer service level. There are three types of *inventory costs*: (1) order costs associated with placing an order, (2) holding costs for carrying inventory until it is sold or used, and (3) penalty costs for unfulfilled customer demand. The *order cost* can consist of fixed cost for each time an order is placed and variable cost for each unit ordered. The *holding cost* is mainly the opportunity cost of the money invested in inventory. But it should represent all inventory carrying costs, including the cost of warehouse space, material handling, insurance and obsolescence. The *penalty cost* is the cost of not having sufficient inventory to meet all customer demands. These shortage costs can be interpreted as the loss of customers' goodwill and the subsequent reluctance to do business with the firm, the cost of delayed or no revenue, and any possible extra administrative costs.

When demand is stochastic, shortages cannot be avoided. A service level is used in a supply chain to measure the performance of such inventory systems. The most common measures of service are (1) α service level, (2) β service level, and (3) γ service level. The first type of service level is an event-oriented criterion. It measures the probability that all customer demands are satisfied within a replenishment cycle. This definition is also called the *cycle service level*, since it measures the fraction of cycles in which a stock out occurs. The β service level, or *fill rate*, is a quantity-oriented measure that represents the fraction of the demand satisfied directly from stock on hand. An example to illustrate the difference between the cycle service level and the fill rate is provided in Table 1.2. The total directly satisfied demand from stock on hand is 12 units, while the total demand is for 15 units. Therefore, the fill rate equals $12/15 = 80\%$. Since the demand exceeds the inventory level in two out of three order cycles, the cycle service level equals 33.3%. The γ service level incorporates the waiting time of the demands backordered. This service performance measure is the fraction of

order cycle	inventory level	demand	satisfied demand
1	4	5	4
2	6	8	6
3	3	2	2

Table 1.2: An example to show the difference between cycle service level and fill rate.

time during which there is no stock out. This service level definition is also called the *ready rate*.

To summarize, an inventory control system is characterized by the number of items, the number of stocking points, the periodicity, the demand nature, the review interval, the order size, and the objective. In this thesis, several models are developed for inventory control systems with different characteristics and replenishment policies. For these models we derive expressions and procedures to compute the average cost and service level based on the definitions as discussed in this section. These basic concepts of inventory theory are used throughout the thesis without much extra elaboration. More details about the scope and contribution of this thesis are discussed in the next section.

1.2 Contribution and scope of this thesis

A lot of real-world problems in inventory management are not addressed in the literature. Silver [266] has recently published a paper in which he summarizes what needs to be done to bridge the gap between theory and practice. One of his main suggestions is to look carefully at characteristics of current practices and develop accurate models that represent the real world without invalid assumptions. As described at the beginning of this chapter, inventory systems have changed. The inventory models should therefore be adapted accordingly. It is much better to obtain good solutions for a realistic model instead of theoretically nice and optimal solutions to a model which is not realistic for practical purposes. Thereby, understandable solution approaches and decision rules are better than optimal solutions that are neither understood nor accepted by management. As mentioned before, the customer focus and the accompanied developments to increase customer satisfaction are essential characteristics that should be included in inventory models. New approaches and solution techniques have to be developed in order to apply inventory theory in real-world inventory control systems. In this perspective, the goal of this thesis is to develop and solve inventory models that reflect current practices in customer behavior and service. The goal of this section is to provide more details on what has already been done in inventory theory, and why new solution methods have to be developed. The contribution of this thesis to the existing models from the literature is twofold: (1) a service level constraint is included next to a cost minimization objective function, and (2) excess demand is not backordered.

Both aspects (service models and no backlogging of excess demand) are investigated in three specific practical settings. These settings are chosen such that they represent all kinds of inventory systems that are encountered in the real world. In the first setting, an inventory problem in after-sales is considered. In the second setting, the influence of lost sales on the replenishment policy is illustrated. In the third setting, a rental company is considered in which customers behave differently in case of a stock out (substitute demand over multiple items and multiple locations, backorders and lost sales). Redistribution of stock is also considered in the third setting to prevent stock outs. The practical importance and the contribution with respect to the existing literature for each setting are described in the remainder of this section.

After-sales services

After-sales services are activities taking place after the purchase of the product and are devoted to support customers in the usage of the product. It includes maintenance, repair, and upgrading. Delivering high levels of customer satisfaction through after-sales activities can increase loyalty, and thereby sales. The strategic role of after-sales services helps a company not only to retain customers but has a direct impact on the image of the company as well. If these services can be offered at a fixed or guaranteed rate, they can be a significant competitive advantage (Gaiardelli et al. [90], Saccania et al. [252]). Companies need to think of a lifetime relationship in terms of both products and customers. The product sale is only a small part of the overall value during the complete product life cycle. It is the start of the customer relationship.

After-sales services become more important when the average unit price of an end product is high and the length of the product life cycle is over 10 years. In order to satisfy customer demand, companies offer customized products next to a variety of standard products. Typical examples are found in an industrial environment. Customization results in unique products and items that would not have been produced otherwise. In general, this tends to increase the number of items that have to be stored for after-sales operations.

Spare part inventory management is different compared to most inventory control systems. It is characterized by a lot of different items with low demand volumes per item, and demand fluctuations. The long life cycles of items make inventory control even more demanding. One subgroup within spare parts are *consumable parts*, which are not economically repairable and are discarded when worn out or broken. *Repairable parts* (or *recoverable parts*) are repaired when they fail. There are a lot of studies about spare part inventories. See Kennedy et al. [159] for an overview. A well-known subclass of spare part inventory models is the METRIC model introduced by Sherbrooke [264]. They include the repair of failed items at a repair shop. METRIC has been refined and extended in several ways (Muckstadt [213]).

In the first setting of this thesis, the focus is on replacing broken items to repair machines. Such repairs are performed on location, where repairmen bring along a set of spare parts to fix the problem. Since it is unknown beforehand which items have to be replaced, there is a probability that not all required items are in this set of spare parts. The machine is only fixed when all required items are available in the right number. The inventory decision how many items to bring along is called the *repair kit problem*. The service performance of this after-sales activity is measured with the probability that all required items are available. This service definition is called the *order fill rate*. Notice the difference between this definition and the β fill rate. The former service definition includes the availability of multiple items, whereas a regular fill rate measures the availability of a single item. Two types of models have been developed in the literature in case of an order-based service measure: cost models and service models. In a *cost model* the total costs consisting of order, holding and penalty costs are minimized. The objective function in a *service model* is to minimize the order and holding costs subject to a service level constraint. Most of the literature addresses the cost model. In practice, however, the service model is preferred. It is often very difficult for management to measure and quantify some of the inventory costs. In particular, the cost associated with *loss of goodwill* due to excess demand is difficult to estimate accurately. Most firms have less difficulty in specifying their desired service level. The cost model and service model are however closely related to each other (Van Houtum and Zijm [293]).

The difficulty in formulating both types of models, is to find an exact expression for the order fill rate. In the models developed in the literature on the repair kit problem, assumptions are made which do not correspond to current practices: replenishments after each customer demand and unit-sized demand per item (an overview on all literature is provided in Chapter 2). These assumptions are relaxed by Teunter [285]. Therefore, this model reflects current practices the best. However, the assumption is made that all required items that are available on stock are always used despite the fact whether the repair can be finished. This is not in accordance with current practices, in which items are only used when all required items are available to complete the repair. Otherwise, the items that are available can be used to fulfill future demands. Unfulfilled repairs are taken care of with a return visit.

The inventory control system is classified as a multi-item inventory model in which excess demand results in emergency transshipments. Based on the special structure of overnight replenishments, the system is characterized by periodic reviews, zero lead times, and no order costs. Consequently, the repair kit problem is modeled as a single-period problem in which the appropriate base-stock levels have to be determined. A more detailed description is provided in Chapter 2. The goal is to develop a more general service and cost model for this inventory control system in which the beforementioned assumptions are relaxed. Consequently, the model incorporates all aspects that are observed in practice. An exact expression for the job fill rate is derived as well as a heuristic procedure to determine near-optimal stock quantities. Therefore, the proposed solution procedure can directly be applied in real life to set the base-stock levels.

Lost sales

Traditionally, retailers are the only party in the supply chain to interact with end customers (see Figure 1.1). Currently, the retail environment is changing. First, the customer behavior has changed. Second, retailers are not the only distribution channel for manufacturers anymore. The first change is a direct result of the globalization of the economy and the use of internet as a distribution channel. The number of products to choose from is enormous. When a specific brand or product is not available, the customer does not wait but looks for a substitute product that meets its price and quality expectations (Ervolina et al. [79]). Besides a weakened customer loyalty, consumers also look for alternative locations to buy the same or a similar product. In both scenarios, the original demand is lost. The second change in the retail environment has to deal with the fact that manufacturers are integrating supply chain processes to become more cost-efficient, flexible, and responsive to customer demands (see, e.g., Wadhwa et al. [302], Hsueh and Chang [120], Zhao et al. [325], Fabbe-Costes and Jahre [82]). Due to the great potential of the internet to sell directly to customers, many brand name manufacturers, including Nike, Cisco Systems, Hewlett-Packard, IBM, and Apple, have added direct channel operations (Tsay and Agrawal [288]). More companies seriously evaluate such strategies. The world's largest trade publisher Random House started an online bookstore in 2005 to sell their own books directly to readers, putting them in direct competition with Barnes & Noble and Amazon.com [1]. Internet has become an important retail channel. According to the US Census Bureau [2], the online retail sales comprised of more than 93 billion USD in 2005, which represents 2.5% of all retail sales excluding travel. They expect that the total online sales is going to increase to 271 billion USD in 2011. Therefore, the retail environment is still growing and becomes more competitive every day. Now more than ever, the behavior of customers has to be taken into consideration.

The majority of the models available in the literature assume backlogging when customer demand cannot be fulfilled immediately with inventory on hand (see Zipkin [327], Silver et al. [268]). The main reason for this development is because Karlin and Scarf [155] proved that the (R, s, S) policy is optimal for periodic review inventory models with a fixed lead time and backorders. Different techniques have been developed to set the reorder level and the order-up-to level for the cost model (Porteus [236], Federgruen and Zipkin [85]) as well as the service model (Tijms and Groenevelt [287]). When there is a positive lead time and excess demand is lost rather than backordered, the optimal policy is extremely complex. The backorder model is therefore used as approximation model for the lost-sales case. From a theoretical point of view it is not a bad idea to start with these relatively simple models. However, with the experiences gained, more realistic models have to be constructed. The retail market has become very competitive and customers are not as loyal anymore to a specific brand or store as they used to be. Therefore, it is not reasonable to assume that customers are willing to wait for the next order delivery when a product is out of stock. Either a retailer has to perform an emergency transshipment, or the customer buys another brand or product, or goes to a different retailer. In all scenarios the original demand is not dealt with as in the normal replenishment process and can therefore be regarded as lost instead of backordered. As shown by Bijvank and Vis [38], the current solution techniques are not sufficient when lost sales is of importance. Therefore, different approaches have to be developed for the lost-sales case.

Most of the research on lost-sales inventory systems describe a continuous review process. See, for example, Hadley and Whitin [105], Hill [114], Johansen and Thorstenson [137], Hill and Johansen [118]. In many real situations the inventory is reviewed periodically at regular intervals. Such systems are more practical in terms of coordinating the replenishments and they offer the opportunity to adjust the reorder level and the order size. This is a desirable property if the demand pattern is changing in time. The periodic review models that do exist in a lost-sales setting only consider a cost objective with no order costs (see, e.g., Zipkin [328], Johansen and Thorstensen [138]). However, as indicated above, the service provided to customers is very important for the competitive position of companies. Therefore, service constraints should also be included in the inventory control model. Furthermore, all papers assume either that the lead time is an *integral multiple* of the review period length or *fractional lead times* (a lead time shorter than the length of a review period). In practice, the lead time is a constant (or the variability is negligible) and mainly determined by the supplier. The lead time should, however, also include the time for a retailer to tranship the delivered items to the shelves. Therefore, the lead time is not strictly related to the review period length in real inventory systems. In order to set the review interval, the retailer has to know the impact of different review period lengths.

The inventory system in this setting is classified as a single-item inventory control system with periodic reviews and lost sales. The goal in this setting is to develop and compare mathematical models to find optimal and near-optimal replenishment policies for such inventory systems. More general models are developed in which the lead time and review period can be of any length, and where both a cost objective and a service level restriction are considered. We extend these models to include order costs. The determination of the exact best values for the reorder level and order size is extremely difficult, and requires an extensive computational effort. Therefore, we also derive heuristic methods to set the values of these decision variables.

Customer behavior to stock outs

In the after-sales setting demand is assumed to be satisfied with an emergency delivery outside the regular replenishment process when there is an out-of-stock situation, whereas excess demand is assumed to be lost in the second setting. There are also other reactions to deal with stock outs (e.g., substitution). Moreover, it is important to understand that it is not possible to assume the same behavior of all customers. Therefore, a customer choice model is studied in the third setting in which excess demand is backordered, lost, or substituted. A substitution can be either a substitution of a different location or a different product.

The inventory control systems in this setting are classified as multi-item, multi-location inventory models. As illustrated by Figure 1.1, inventories are kept at multiple locations in a supply chain. Each stage or location in the supply chain is called an *echelon*. When these stages are coupled to each other, the inventory system is called a *multi-echelon* inventory system. For example, many companies use an inventory system with a central warehouse close to the production facility and a number of local stocking points close to the end customers when products are distributed over a large geographical area. It is very difficult to allocate safety stocks optimally in a multi-echelon system. The best-known technique to do this is presented by Clark and Scarf [61]. See also Eppen and Schrage [77], Federgruen and Zipkin [86],[87], Van Houtum and Zijm [292], Federgruen [83], Van Houtum et al. [291], and Verrijdt and De Kok [298].

The most common assumption in multi-echelon models is that shipments among retailers are not allowed. In real life, however, when a retailer cannot satisfy demand directly from stock on hand, it is taken from an adjacent retailer that has the item available on stock. This is a direct result of the behavioral changes of customers, who are not willing to wait anymore for backorders (as discussed in the previous setting). Models which allow the transfer of a product among locations at the same echelon level are called *lateral transshipment* models.

In this final setting we focus on a rental company where lateral transshipments occur with zero lead time in case of emergencies to satisfy customer demand.

Another option is to offer a substitute product when an item is not on the shelves. The goal in this setting is to define a model and set the inventory levels for each item at each location when a service model is considered for an inventory control system with backorders, lost sales, substitutions, and lateral transshipments. There exists no literature in which all the alternatives to a stock out are included.

Table 1.3 summarizes the different characteristics of the inventory control system for each setting. It clearly illustrates that all kinds of inventory systems are considered in this thesis that are relevant for practical purposes. More details on how to read this thesis can be found in the next section.

characteristic	setting 1: after sales	setting 2: lost sales	setting 3: rental company
number of items	multi-item	single-item	multi-item
location	single-location	single-location	multi-location
periodicity	single-period	multi-period	multi-period
demand	stochastic	stochastic	stochastic
review interval	periodic review	periodic review	periodic review
order size	variable	variable, fixed	variable
objective	cost, service	cost, service	service

Table 1.3: An overview of the characteristics of the inventory control systems considered in this thesis.

1.3 Outline of this thesis

Each setting, as described in the previous section, corresponds to a part of this thesis. Each part is written as a stand alone resource and can be read independently from the other parts of this thesis.

In part I, a specific area of after-sales is considered. Namely, the inventory management in field services where the order fill rate is used as service performance. This is also called the repair kit problem. Chapter 2 introduces this problem in more detail and provides an overview of the literature on order fill rates. A mathematical model and solution approach are developed in Chapter 3. In this chapter, numerical results and a case study show the performance of the solution approaches for the cost and service model. This part is based on research in Bijvank [34] and Bijvank et al. [40]

Part II addresses a single-item inventory system with lost sales. A literature overview on lost-sales models is provided in Chapter 4. Since periodic reviews are common in practice, we focus on such replenishment policies in the remainder of part II. In Chapter 5 a general model is developed to compare replenishment policies when no fixed order costs are charged for each order. Consequently, base-stock policies are commonly used under such circumstances. A new type of policy is proposed which restricts the order size to a maximum. Chapter 6 extends the model of Chapter 5 for the case of fixed order costs, and also performs a comparison between different replenishment policies. A case study is discussed in Chapter 7 for a hospital setting, which is characterized by fractional lead times and limited available capacity. Part II is based on research in Bijvank and Johansen [36], Bijvank et al. [35], and Bijvank and Vis [37].

In part III, substitution and lateral transshipments are considered as alternatives to excess demand besides backlogging and lost sales. To be more specific, we consider a rental company in which customers visit multiple locations. The introduction to this setting and the relevant literature is discussed in Chapter 8. We develop an inventory control model with periodic reviews in Chapter 9 to compute the service level perceived by customers. This model is used to determine the appropriate number of products to keep on stock at each location. The work in this part is based on research in Vis et al. [300] and Bijvank et al. [39]. In Chapter 10 conclusions and suggestions for further research are presented.

Part I

After-sales services

Chapter 2

Order-based service levels

Production and service environments rely upon high levels of automation. A breakdown of a (bottleneck) machine often results in a breakdown of the entire process, which affects the production lead time and ultimately the complete supply chain. To ensure the optimal availability of machines, customers expect to get more than just the physical product when they buy a machine or piece of equipment. Namely, they also expect to get services regarding any malfunctioning of the machine during its life cycle. The activities regarding repair, maintenance, and upgrades of products are called *after-sales services*. Such activities are commonly found in industrial environments. However, there is also a growing need for such product support in a service environment. For example, the repair of copiers, coffee-machines, and computer systems. Offering a good after-sales service is an important marketing and differentiation strategy among competitors.

One particular type of after-sales service is a local repair service at the customer's facility. This is called *field service*. A repair is mostly characterized by the replacement of multiple items. The related service level definition for a repair is not as straightforward as described in Section 1.1 on the basics of inventory management. A customer is only satisfied whenever the failed machine is fixed. The customer service is defined as the probability that all items are on stock to finish the repair. This service measure is called the *job fill rate* or the *order fill rate*. To our knowledge, there are two inventory systems in the literature which consider the job fill rate as service perceived by customers. The first system is the inventory control problem of spare parts for the field service as described above. This problem is called the *repair kit problem.* The second system is an *assemble-to-order* (ATO) system in which customer requests are assembled out of multiple components. In both systems, the customer's order consists of a set of different items for which an order-based service level is essential. The goal of this chapter is to illustrate that there is lack of literature on inventory models which consider an order fill rate, while it is an important characteristic in real-world inventory systems. Section 2.1 provides a more detailed description of the two areas in which order fill rates are found. A literature overview on the order fill rate is addressed in Section 2.2. Our contribution is outlined in Section 2.3.

2.1 Problem description

Companies often use an item approach to determine the reorder level and order size. In an item approach, these decision variables are set for each individual item independent of other items. This is not appropriate when a customer's demand or order consists of several different items in different numbers. As a result, the order-based performance can be very poor while the item-based performance measures are satisfactory. As illustrated by Mamer and Smith [195] and Song [275], the product of item fill rates over all items provides a lower bound on the order fill rate. Therefore, the order fill rate should be used as service measure in systems where multiple items are requested by a single customer. Smith et al. [274] introduced the concept of order fill rates in a context with repairs (the repair kit problem). Such service performance measures are also considered in assemble-to-order (ATO) systems. Both inventory problems are described in this section.

The repair kit problem

If a machine failure occurs, the supplier of the machine is contacted to perform a repair. A technician with the right skills to maintain the machine is scheduled to visit the customer. Each morning, the technicians receive a list of these call points and they travel around to repair the broken machines. Since it is not known in advance which parts of the machine have to be replaced, the repair person takes along a selection of the spare parts in the car. This set of parts is referred to as the *repair kit.* The technician can only complete a repair if all required spare parts are available in the right quantity in the repair kit. When one or more parts are missing, the technician cannot fix the machine and has to return when the car is restocked with all the items that are required to finish the repair. This extra visit is called a *return-to-fit* (RTF) visit. An important logistics decision problem for technicians is to determine which spare parts to put in the repair kit and in which quantities to avoid return-to-fit visits. This problem is called the *repair kit problem.*

According to experts in the field (see Bijvank [34] and the case study in Section 3.4), companies usually base the contents of their repair kits upon experiences and practical limitations (e.g., the capacity of the car and the amount of money spent to purchase the contents of the repair kit). It would be more efficient to have a systematic procedure to determine the contents of a repair kit. Such a procedure could be based on costs but also on a service level granted to customers. We distinguish between two types of costs, namely holding costs and return-to-fit (RTF) costs. A fixed amount of *holding costs* is incurred for each unit that is stored in the repair kit of the car (see also Section 1.1). *RTF costs* are involved when a technician has to return because at least one of the required parts is not available in the repair kit. These RTF costs usually consist of the actual labor and driving costs, as well as costs due to loss of goodwill.

To compute the expected RTF costs or the service perceived by customers, the job fill rate has to be calculated. It is, however, difficult to find an exact mathematical expression for the job fill rate. First, because it depends on the availability of all required spare parts. Second, because this availability depends on the number of repairs already performed with the repair kit. A second repair has a higher probability to result in an RTF visit compared to a first job since fewer items remain in the repair kit after the first job. Therefore, the concept of a *tour* is introduced in which a sequence of jobs is performed before the repair kit is restocked. The number of jobs performed between two restock moments is called the *tour size.* In practice, cars are restocked on a daily basis during the night. There are also practical examples where cars are restocked after two days of work (see Heeremans and Gelders [111]). Normally a car is replenished at or from a central depot, which has enough spare parts available to restock all cars immediately. Figure 2.1 depicts this replenishment process. Since there is no repair performed between the order placement at the central depot after a tour and the order delivery in the car the next morning, the repair kit problem can be categorized as a periodic review inventory system with no lead times. Consequently, the repair kit problem can be represented by a single-period model (see also Section 1.1). Since the service is measured in terms of the order fill rate, it requires a multi-item approach. Excess demand is dealt with by means of RTF visits. In case of a return visit, the technician already knows which items are required. A repair person puts these items in the repair kit in the right quantity such that the repair can be completed during the next visit. Thus, an RTF visit is seen as an emergency transshipment and is dealt with outside the normal replenishment process of repair kits.

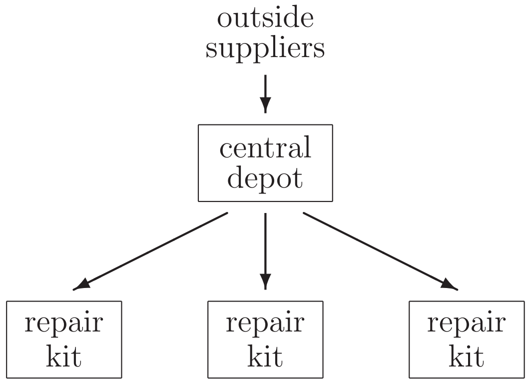


Figure 2.1: The replenishment process for repair kits.

Assemble-to-order systems

In *assemble-to-order* (ATO) manufacturing systems inventories are kept only at the component level. Final products are assembled in response to customer orders. In ATO systems the customer's demand is for one product type. This is called the *type of demand*. Each type of demand corresponds to the demand of a subset of components that constructs the product type. In pure assembly systems, there is only one type of demand (or one end product). As described in Chapter 1, the number of products that a customer can choose from is growing. Therefore, systems with multiple types of customer demand are generally found in practice. This is also our focus.

An order can only be assembled or processed when all items (or components) are available. When a customer demand cannot be fulfilled directly, the missing components are produced and backordered. The optimal inventory control policy for this system is unknown. Base-stock policies are widely adopted in practice. This means that an order for the production of a new item is placed when it is taken from the on-hand inventory. If no stock is available, the item is backordered according to the same production process. This resupply process of items can occur with continuous or periodic reviews. Since the time to produce a single unit (i.e., the *lead time*) is not zero, this inventory system is classified as a multi-period model for which the base-stock levels have to be determined (see also Section 1.1). See Song and Zipkin [279] for an overview on all types of ATO systems.

Table 2.1 gives a summary of the inventory control characteristics for the repair kit problem and ATO systems. At first sight these inventory systems are completely different. Consequently, there has been a separate development of models in the literature for such inventory systems. However, both systems determine base-stock levels based on an order-based service measure. Finding an exact expression for the order fill rate is not trivial. The different types of assumptions and inventory system characteristics to develop computational procedures for the order fill rate are discussed in the next section.

repair kit problem	ATO systems
- order-based service measure	- order-based service measure
- base-stock policy	- base-stock policy
- periodic review	- (mostly) continuous review
- (in)dependent item demand	- dependent item demand
- excess demand is lost	- excess demand is backlogged
- zero lead times	- positive lead times
- single-period problem	- multi-period problem

Table 2.1: A comparison of the inventory control characteristics for the repair kit problem and ATO systems.

2.2 Literature overview

As indicated in the previous section, there is a clear distinction between the models for the repair kit problem and ATO systems. To give a complete literature overview on order-based service measures, the literature on both inventory systems is discussed separately in this section.

The repair kit problem

In the literature on the repair kit problem, there is a distinction between cost models and service models. In a *cost model*, the holding and RTF costs are minimized to make the trade-off between holding costs and RTF visits, whereas in a *service model* holding costs are minimized subject to a service level constraint. This latter model is preferred in practice due to the difficulty to quantify the extra cost for an RTF visit. Moreover, this type of model explicitly incorporates a customer service level criterion such that a minimum quality of service is guaranteed (see also Section 1.2). However, both models are clearly related since a higher probability for an RTF to occur results in lower customer service and higher expected RTF costs.

The main difficulty in formulating both types of models is to find an exact expression for the probability that an RTF visit will occur. Therefore, papers that address the repair kit problem impose several assumptions which are not realistic in many practical situations. The first models that have been developed assume that the technician returns to the central depot or warehouse after each job to restock the repair kit. Consequently, each job has the same probability of being completed during the first visit. Another assumption made in these models is that at most one unit of each spare part can be used for the repair. When both assumptions hold, at most one unit of each part type is added to the repair kit to obtain optimal contents. Smith et al. [274] develop a cost model under these assumptions and formulate a binary integer program. In their solution technique to find the optimal contents of the repair kit, the part types are ranked by the ratios of holding cost to usage frequency. The contents of $N + 1$ repair kits is determined, where N is the number of part types. Each repair kit M_k contains the first k part types according to the ranking. The optimal solution is found by calculating the total expected costs for each of the repair kits. Optimality is guaranteed since this ranking selects the most preferred part types first. Under the same assumptions Graves [99] develops a service model and transforms it in a binary knapsack problem. A greedy heuristic procedure is proposed to add part types one by one to the repair kit until the service level constraint is satisfied. Hausman [108] proposes a randomized-strategy approach for this service model. A similar greedy heuristic is proposed by Cohen et al. [63] in which more than one unit of any part type can be required to complete a repair. They also derive a duality heuristic based on Lagrange multipliers to solve the service model. Schaefer [260] extends the service and cost model to the case where failed parts are repaired at a repair facility. Multiple units can be kept on inventory in this setting because of the repair facility.

So far, all authors assume independence between the different part type failures causing the breakdown. Mamer and Smith [195] relax this assumption of independence between the failure probabilities by defining a representative collection of job types where each job type corresponds to a set of demands for parts. They formulate the problem as a network problem and solve it with a max flow/min cut algorithm. Figure 2.2 shows an example of the network formulation in which each job type consists of one or more part type demands. This formulation allows for a job to require more than one unit of each part type. In that case, an extra part node has to be added to the graphical representation of Figure 2.2 for each extra unit that is demanded of the same part type.

As mentioned before, it is difficult for management to set the penalty cost for

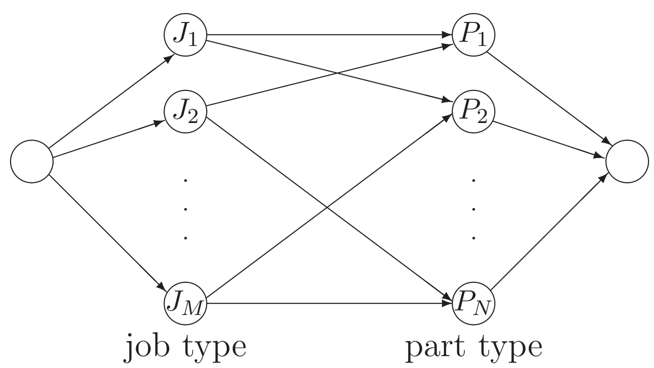


Figure 2.2: Network representation of the repair kit problem.

an RTF visit. In addition to the cost model of Mamer and Smith [195], March and Scudder [197] determine a set of optimal policies and a range on the corresponding penalty cost for which each policy is optimal. They show that as the penalty cost increases, the optimal policy always includes all the particular part types that were stocked at lower penalty costs. This result implies that standard exchange curve techniques can be used to find an optimal solution. A similar result is obtained by Mamer and Smith [196] and Brumelle and Granot [46]. Mamer and Smith [196] also add the extra option to replace an entire machine when a spare part is not available. An analogy between the network formulation of Mamer and Smith [195] and a problem where projects have to be selected with limited resources is presented by Mamer and Shogan [194].

At the end of their paper, Mamer and Smith [195] remark that a job (or repair) can be interpreted as a tour which might involve more than one repair. The penalty cost would correspond to the cost of having at least one RTF visit in a tour instead of the cost for one RTF visit. Heermans and Gelders [111] explicitly model tours of a fixed size. They also include a space limitation and formulate the service model as an integer linear program. The space constraint is dropped when they solve the model with a similar kind of knapsack heuristic as Graves [99]. As pointed out by Teunter [285], they express the expected service level as the probability of not having an RTF visit in a tour. This does not correspond to the order fill rate definition, and he defines it as a *tour fill rate*. When tours can be of any fixed size, Teunter [285] gives an exact expression for the order fill rate under the assumption that at most one unit of each part type is used in a repair. He reformulates the knapsack heuristic as proposed by Heeremans and Gelders [111] to solve the cost model. For the case in which multiple units of the same part type may be needed, the author proposes a second heuristic in which the order fill rate is approximated based on part fill rates. The use of both heuristics is also illustrated when part failures are dependent.

Table 2.2 provides an overview of all assumptions made in the literature on the repair kit problem as discussed above. The second column specifies the assumption that the demand for an item is unit-sized. In the second assumption, the repair kit is restocked after each customer demand when the tour size equals one. The third assumption indicates whether the failure probabilities for the parts are dependent (D) or independent (I). Besides the assumptions, Table 2.2 also summarizes which type of objective function is used (C represents the cost model and S the service model), whether an exact or approximated expression is formulated for the order fill rate, and which type of solution approach is proposed.

	demand size	tour size	part failure	objective	order fill rate	solution approach
Smith et al. [274]	1	1	I	C	exact	optimal
Graves [99]	1	1	I	S	exact	heuristic
Hausman [108]	1	1	I	S	exact	optimal
Cohen et al. [63]	> 1	1	I	S/C	exact	heuristic
Mamer and Smith [195]	> 1	1	D	C	exact	optimal
March and Scudder [197]	> 1	1	D	C	exact	optimal
Mamer and Smith [196]	> 1	1	D	C	exact	optimal
Brumelle and Granot [46]	> 1	1	D	C	exact	optimal
Mamer and Shogan [194]	> 1	1	D	C	exact	optimal
Heeremans and Gelders [111]	> 1	> 1	I	S	approx	heuristic
Teunter [285]	1	> 1	D	C	exact	heuristic
Teunter [285]	> 1	> 1	D	C	approx	heuristic

Table 2.2: The assumptions made in the literature on the repair kit problem: single or multi-unit demand, single or multiple repairs in a tour, dependent (D) or independent (I) part failures in a cost (C) or service (S) model.

Assemble-to-order system

In most literature on ATO systems it is common to assume a Poisson process to represent the demand process for each product type. The customer arrival process for a type of demand is assumed to be independent of the other demand types. Consequently, the demand for each item is a compound Poisson process whose rate is the sum of that of the individual demand processes. Notice that the commonality of components in ATO systems is similar to the definition of part failure dependencies in the network formulation of the repair kit problem in Figure 2.2. The job types correspond to the demand types in an ATO system and the part types to the components. Furthermore, the assumption of a first-come first-serve (FCFS) policy is used in the supply process for backorders. The available literature on ATO systems addresses different performance indicators. For example, the average number of orders that are not completely filled (also referred to as the average *order-based backorders*) or the time delay in customer orders. Also different assumptions are made about the supply process, as is described in the remainder of this section.

Cheung and Hausman [55] claim to be the first authors to address a model with a service level based on the demand for multiple items in an inventory system with continuous reviews. They consider their model in a repair shop context. However, Schaefer [260] has also developed a model for repairable items in which order-based performance measures are considered under continuous review. The author assumes that at most S_i units are stocked for item i. When an item fails it is replaced and the failed item is repaired. When all S_i units are in repair, the demand for item i is lost (or dealt with as an emergency transshipment as described for the repair kit problem). Consequently at most S_i units can be in repair simultaneously. Vliegen and Van Houtum [301] consider a similar inventory system, where the number of units in production (or repair) is also limited to S_i and there is no buffer for queueing (i.e., excess demand is lost). They, however, assume dependencies in the demand for items and in the returns of repaired items. In the model by Cheung and Hausman [55] there is no limitation on the number of items in repair. They assume backordering when all S_i units are in repair. Another assumption underlying their model is complete and instantaneous cannibalization to derive the distribution function and the expectation of the number of orders backlogged.

The cannibalization assumption is not realistic in ATO systems. Hausman et al. [109] relax this assumption for a system with constant replenishment lead times. They compute the order fill rate within a pre-specified time interval for a periodic review model in which the demand during a period has a multivariate normal distribution. The base-stock levels are determined such that the number of orders that are fulfilled within a pre-specified time limit is maximized for a given inventory investment budget. Cheng et al. [53] develop a model in which the expected inventory costs are minimized subject to an order fill rate constraint. Zhang [323] also solves this service model with periodic reviews. The author assumes that products with a higher priority receive backlogged components before products with a lower priority instead of the FCFS policy. A combination of a FCFS policy and a fair-share allocation rule is used by Agrawal and Cohen [5].

Exact results for the order fill rate in a continuous review model with constant lead times are presented by Song [275]. A more general replenishment policy is considered in Song [276], in which an order of size nQ is placed when the inventory position of an item falls to or below the reorder level where n is the smallest integer so that the inventory position after ordering is above the reorder level. When $Q = 1$, the policy reduces to a base-stock policy. The average order-based backorders is used as performance measure by Song [277].

Zhang [324] assumes random lead times to find an expression for the expected waiting time. The supply process is performed on a single-server machine with an infinite buffer queue. Song et al. [278] derive expressions for the exact waiting time distribution and order fill rate in case of exponentially distributed lead times in which the backlog queue is finite. Their procedure is computationally complex. Exact performance calculations can only be obtained for small to medium-sized systems. Dayanik et al. [69] develop easier-to-compute performance estimates to overcome the complexity of the exact approach. The trade-off between inventory levels and the lead time is studied by Glasserman and Wang [95] for a more general demand structure assuming high fill rates. Iravani et al. [126] model the supply of components as a single-server machine with batch productions.

In almost all papers mentioned above, each item is produced on a single server when the production time (or lead time) is stochastic. Lu et al. [188, 189] use an infinite-server queue as supply process. They generalize the unit-sized demand model of Cheung and Hausman [55], in which at most one unit of each item is used for each demand type. The same objective function is used in both papers in which the average order-based backorders is minimized subject to a budget constraint. Lu and Song [189] formulate a cost model in which order-based backorder costs are included to determine optimal base-stock levels. Lu [187] extends the model of Lu et al. [188] by relaxing the assumption of a Poisson arrival process for each demand type. The author allows the arrival process to be a general renewal process.

An overview of the different assumptions made in ATO systems is provided in Table 2.3. The second column indicates whether the replenishment process is continuous (C) or periodic (P). The second assumption indicates whether the demand size for each part type is unit sized (1) or multiple units can be required (> 1). In the fourth column, the number of machines for the supply process of each part type can be a single machine (1), a fixed number of machines (S_i) or an infinite number of machines (∞). The lead time distribution for this supply process can be deterministic (D), exponential (E) or general (G).

Based on this literature overview, it can be concluded that the amount of research on ATO systems is more extensive compared to the repair kit problem. More literature is found for ATO systems under different characteristics and assumptions, as illustrated in Table 2.3. Therefore, we focus on the repair kit problem in the remainder of this part of the thesis.

2.3 Contribution and outline of part I

A lot of different assumptions are made in the inventory models of Section 2.2 to find an exact expression for the order fill rate. In almost all papers mentioned in the previous section, the authors assume that when a repair or order cannot be completed due to the unavailability of one or more required parts, the subset of required parts that is available is kept aside as committed inventory and cannot be used in the following jobs. This seems logical to assume in an ATO system in which lead times are involved. However, this does not make sense to assume for the repair kit problem, since it increases the probability for an RTF visit to occur. Consequently, the models developed for the repair kit problem are not generally applicable in real life. Therefore, we develop a model for the repair kit problem in which the assumption of committed inventory is relaxed when a customer demand

	review interval	demand size	no. servers per part	lead time
Hausman et al. [109]	P	> 1	∞	D
Cheng et al. [53]	P	> 1	∞	G
Zhang [323]	P	> 1	∞	D
Agrawal and Cohen [5]	P	> 1	∞	D
Schaefer [260]	C	1	S_i	E
Vliegen and Van Houtum [301]	C	1	S_i	D/E
Cheung and Hausman [55]	C	1	∞	G
Song [275][276][277]	C	> 1	∞	D
Zhang [324]	C	1	1	G
Song et al. [278]	C	1	1	E
Dayanik et al. [69]	C	1	1	E
Glasserman and Wang [95]	C	> 1	1	E
Iravani et al. [126]	C	1	1	E
Lu et al. [188][189]	C	> 1	∞	G
Lu and Song [189]	C	1	∞	G
Lu [187]	C	> 1	∞	G

Table 2.3: The assumptions made in the literature on ATO systems: continuous (C) or periodic (P) reviews, single or multi-unit demand, the number of servers available for the supply process, in which the lead time distribution can be deterministic (D), exponential (E) or general (G).

is not satisfied immediately.

In order to relax this assumption, the entire demand has to be cancelled when one of the required items is not available on stock. Such a characteristic has already been addressed in ATO systems with a finite queue length in case of backorders. In this respect, a distinction is made between total order service models and partial order service models. In a *total order service* (TOS) model the entire order is lost when the queue for one of its backlogged items is full. In a *partial order service* (POS) model only the backlogged items which face a full queue are lost. When the buffer for all queues is zero in a TOS model, an entire order is lost when at least one of the required items is out of stock. This property reflects current practices for the repair kit problem. However, it is only of importance when the repair kit model incorporates a tour which consists of more than one job. Otherwise, the repair kit is restocked immediately after a job. As discussed in Section 2.2, most models assume a tour of size one in the repair kit problem. Teunter [285] is the only author to find an exact expression for the order fill rate in a model with a tour consisting of multiple jobs. However, the assumption is made that at most one unit of each part type can be used in a repair. This is not observed in practice. Therefore, we develop a more general model for the repair kit problem in which multiple units of each item can be used for a repair and the tour size is a random variable. An exact expression for the order fill rate is derived in Chapter 3, and a solution method is proposed for the cost and service model to deal with this general structure.

Chapter 3

The repair kit problem

When a customer's demand consists of multiple items, (s)he is only satisfied when all requested items are available. Therefore, the performance of such inventory systems should be defined by an order-based measure instead of an item-based measure. In the previous chapter, an overview of the literature on models with such an order-based service measure was presented. In this context, the repair kit problem and assemble-to-order (ATO) systems have been introduced. Different inventory system characteristics are considered and a number of assumptions are made in order to calculate the order fill rate. All models developed in the literature assume that items that are available to satisfy a customer request are taken from the shelves, even when the demand for one (or more) item(s) exceeds the inventory level(s). This does not correspond to current practices of field services, where available items of unfinished jobs are usually used in the following jobs whenever required. The goal of this chapter is twofold. First, a cost and service model is developed in Section 3.1 for the repair kit problem in a *general setting*, where multiple units of each item can be required for a job, multiple jobs are performed in a tour and no items are taken from the repair kit when a repair is not finished. Hardly any assumptions are imposed on the models, such that the models are generally applicable in real life. To the best of the author's knowledge, this is the first research to derive an expression for the job fill rate where the assumption of committed inventory is relaxed when a repair cannot be completed during the first visit due to an out-of-stock occurrence of a requested item. The second goal is to propose new solution procedures in Section 3.2 to determine the contents of the repair kit based on both models which incorporate this new expression for the job fill rate. The performance of this technique is tested with extensive numerical results in Section 3.3 and with a case study in Section 3.4. Conclusions and suggestions for future research are presented in Section 3.5.

3.1 Model

The service and cost model use the same expression for the job fill rate in the repair kit problem. The goal of this section is to derive this expression for the repair kit problem in a general setting where items are not kept aside when a job cannot be completed due to lack of required items and no assumptions are made on either the customer demand or the tour size. The definitions and notation to formulate the cost and service model are introduced first. Next, the exact formulation of the job fill rate is derived for the repair kit problem in the general setting.

Definitions and notation

As explained in Chapter 2, the repair kit problem deals with the selection of units of different part types that are put in the car of a technician. The same notation is used as in Teunter [285] to model the repair kit problem. A repair kit is denoted by S, where n_i represents the number of units of part type i in repair kit S after it is restocked. The number of different part types that are considered to put in the repair kit is denoted by N, so $S = [n_1, \dots, n_N]$. As explained in Section 2.2, the objective function in a cost model is to minimize the expected total costs consisting of the holding and RTF costs. Calculating the total holding costs is trivial, since each part type i has its own fixed amount of holding cost H_i per tour (i.e., per replenishment cycle). The total holding costs of repair kit S are denoted by $C_H(S) = \sum_i n_i H_i$. As stated in Section 2.2, the expected total RTF costs are related to the expected job fill rate.

The tour size is a stochastic random variable denoted by M with a probability distribution function $P(M = m)$ and average $E[M]$. The maximum number of jobs that can be performed in a tour equals $\overline{M}$. For a given repair kit S and tour size m, the expected job fill rate is given by $\gamma^{job}(S, m)$. The expected job fill rate of a repair kit S equals

$$\gamma^{job}(S) = \sum_{m=1}^{\overline{M}} P(M = m) m \gamma^{job}(S, m) \Big/ E[M]. \tag{3.1}$$

The expected number of RTF visits equals the total expected number of jobs in a tour minus the expected number of completed jobs in a tour. Hence, the expected

RTF costs equal

$$\begin{aligned} C_{RTF}(S) &= P_{RTF} \sum_{m=1}^{\overline{M}} P(M=m)m\left(1-\gamma^{job}(S,m)\right) \\ &= P_{RTF}E[M]\left(1-\gamma^{job}(S)\right), \end{aligned} \tag{3.2}$$

where P_{RTF} denotes the penalty cost for a return-to-fit visit. In Equation (3.2), the penalty cost is multiplied with the expected number of RTF visits in a tour with size m and the probability for this to occur. The *cost model* can be formulated as

$$\begin{array}{ll} \text{minimize} & C_{RTF}(S) + C_H(S) \\ \text{subject to} & n_i \geq 0, \end{array} \tag{3.3}$$

and the *service model* as

$$\begin{array}{ll} \text{minimize} & C_H(S) \\ \text{subject to} & \gamma^{job}(S) \geq \beta \\ & n_i \geq 0. \end{array} \tag{3.4}$$

A general demand process is considered, where $p_i^{job}(j)$ represents the probability of requiring j units of part type i to perform a job. At most L_i^{max} units of part type i are required in one job.

Order fill rate

Next, a closed-form expression for $\gamma^{job}(S,m)$ is derived where L_i^{max} and $\overline{M}$ can be any number. Teunter [285] gives a closed-form expression for the expected job fill rate with a fixed tour size $\overline{M}$ (i.e., $P(M=\overline{M})=1$) and unit-sized demand for each item (i.e., $L_i^{max}=1$). Consequently, $p_i^{job}(0)+p_i^{job}(1)=1$. For each of the $\overline{M}$ jobs the author calculates the expected probability to successfully repair the machine. The average job fill rate is then found by adding these probabilities and dividing the sum by $\overline{M}$. The expected probability to have enough units available in the m-th job for part type i depends upon the usage of that part type in the previous $m-1$ jobs. At least one unit should be available for each of the required part types after $m-1$ jobs to complete the m-th job. The probability to use l units of a particular part type in $m-1$ repairs equals the probability to replace that part type in l out of the $m-1$ jobs. This latter is true, because at most one unit is used in one job. This probability follows a binomial distribution function.

The assumption of a fixed tour size is relaxed by conditioning on the tour size q ($1 \leq q \leq \overline{M}$). Consequently,

$$\gamma^{job}(S,q) = \frac{1}{q}\sum_{m=1}^{q}\prod_{i=1}^{N}\left\{(1-p_i) + p_i \sum_{l=0}^{\min\{n_i-1,m-1\}}\left[\binom{m-1}{l} p_i^l (1-p_i)^{m-1-l}\right]\right\}, \tag{3.5}$$

where $p_i = p_i^{job}(1)$ and $1 - p_i = p_i^{job}(0)$. This equation can be substituted into Equation (3.1) to find the expected job fill rate. Notice that this expression assumes that a required part type is always removed from the repair kit, even if the job cannot be completed due to lack of other required parts. In the remainder of this section, we correct for this and relax the unit-sized demand assumption.

The binomial distribution of Equation (3.5) cannot be used anymore when more than one unit of a particular part type can be used in a single job. Therefore, a probability distribution function has to be formulated to express the probability that l units of part type i are available at the beginning of the m-th job. This expression should take the possibility into account that not enough units of a particular part type were available in the repair kit to complete a job before the m-th job, but enough units of the same part type are available to perform the m-th job. Take for instance a situation in which 3 units of part type A are required in the first job, but only 2 units are initially available in the repair kit. This will result in a return visit for this first job. During the second job only 2 units of this part type are required. Since no items are taken from the repair kit at the first job, the second job can be completed. Consequently, a stochastic variable N_i^m is defined as the number of units for part type i that are available in the repair kit to perform the m-th job. An expression for the probability distribution function of this random variable should be derived. This can be done by conditioning on the number of completed jobs V^m out of m jobs and the number of units used during these jobs. Let this latter variable be represented by U_i^r for part type i when r jobs are completed. When k units are used in r jobs that are completed, then $n_i - k$ units are left to perform the m-th job. When $m = 1$,

$$P(N_i^1 = l|V^0 = r) = \begin{cases} 1, & \text{if } l = n_i \text{ and } r = 0 \\ 0, & \text{otherwise.} \end{cases}$$

and when $m > 1$,

$$P(N_i^m = l|V^{m-1} = r) = P(U_i^r = n_i - l|T_i^r = n_i), \qquad \text{if } l \leq n_i, r < m,$$

where the probability distribution function of U_i^r depends on the number of items remaining in the repair kit to perform the r completed jobs (denoted by T_i^r). For example, if $L_i^{max} = 3$ and $n_i = 2$, then a job can only be completed if at most two units of item i are demand (or used). Consequently, U_i^r is only defined for 0, 1, and 2. When more than 2 units are demanded, the job cannot be completed and is therefore not included in U_i^r. Therefore, I only consider the conditional probabilities $P(U_i^r = u|T_i^r = j)$ for $u \leq j$ and $j \leq n_i$. Given the fact that a job can only be completed when all required items are available, we know for sure that the number of items demanded is also used in jobs that are completed and not more units are demanded than available (otherwise the job cannot be completed). Consequently,

$$P(U_i^r = u|T_i^r = j) = \begin{cases} \dfrac{p_i^{job}(u)}{\sum_{k=0}^{\min\{j,L_i^{max}\}} p_i^{job}(k)}, & \text{if } r = 1, u \leq j, \\ \sum_{l=0}^{\min\{L_i^{\max},u\}} \dfrac{p_i^{job}(l)}{\sum_{k=0}^{\min\{j,L^i_{max}\}} p_i^{job}(k)} P(U_i^{r-1} = u - l|T_i^{r-1} = j - l), & \text{if } r > 1, u \leq j, \\ 1, & \text{if } r = 0, u = 0, \\ 0, & \text{otherwise.} \end{cases}$$

We divide by $\sum_k p_i^{job}(k)$ to normalize the distribution function such that $\sum_u P(U_i^r = u|T_i^r = j) = 1$. Next, the probability distribution function for the number of completed jobs out of m jobs has to be specified, which is denoted by V^m. First, let us define $\gamma(m)$ as the probability of completing the m-th job and $\gamma(m|V^{m-1} = r)$ as the probability of completing the m-th job when r ($< m$) jobs have already been completed. The latter probability depends on the number of units requested for each part type and the availability of these units,

$$\gamma(m|V^{m-1} = r) = \prod_{i=1}^{N} \left\{ \sum_{j=0}^{L_i^{max}} p_i^{job}(j) \left[\sum_{l=j}^{n_i} P(N_i^m = l|V^{m-1} = r) \right] \right\},$$

and $\gamma(m) = \sum_{r<m} \gamma(m|V^{m-1} = r)P(V^{m-1} = r)$. Since the m-th job can either be completed or not,

$$P(V^m = r) = \begin{cases} 1, & \text{if } m = 0, r = 0, \\ 1 - \gamma(1|0), & \text{if } m = 1, r = 0, \\ \gamma(1|0), & \text{if } m = 1, r = 1, \\ P(V^{m-1} = r)[1 - \gamma(m|r)], & \text{if } m > 1, r = 0, \\ P(V^{m-1} = r)[1 - \gamma(m|r)] \\ \quad + P(V^{m-1} = r - 1)\gamma(m|r - 1), & \text{if } m > 1, 0 < r \leq m, \\ 0, & \text{if } r > m. \end{cases}$$

To calculate the job fill rate for a given repair kit S and tour size q, the probabilities to finish each of the q jobs are added and divided by q, similar to Equation (3.5),

$$\gamma^{job}(S,q) = \frac{1}{q}\sum_{m=1}^{q}\gamma(m). \tag{3.6}$$

The mathematical formulation of the service model and the cost model is finished when Equation (3.6) is substituted in Equation (3.1) and put in Equation (3.3) and Equation (3.4), respectively.

To illustrate the computation of the job fill rate, consider the following example: A tour has a maximum size of $\overline{M} = 2$, where $P(M = 1) = 1/4$ and $P(M = 2) = 3/4$. The holding cost and usage of the part types is given in Table 3.1. The same example will be used in the next section as well.

i	H_i	$L_i^{\max}$	$p_i^{job}(0)$	$p_i^{job}(1)$	$p_i^{job}(2)$	$p_i^{job}(3)$
part 1	1	1	0.9	0.1		
part 2	2	2	0.8	0.05	0.15	
part 3	6	3	0.7	0.1	0.05	0.15

Table 3.1: Information on the various part types as used in the example for the repair kit problem.

For a repair kit $S = [1, 2, 3]$ the job fill rate equals 95.12% whereas the approximation of Teunter [285] results in a service of 94.99%. In case there is a service constraint of 95%, more units are added to the repair kit based on the approximation. A new algorithm to find the contents of the repair kit based on the service and cost model is proposed in the next section.

3.2 Solution procedure

In this section, an algorithm is developed to solve the service and cost model as formulated in the previous section. First the solution procedure for the service model is presented. The procedure for the cost model consists of the same steps with only a few minor adjustments, which will be discussed as well.

Service model

From Equation (3.4), it can be noticed that the service model looks like a knapsack problem. Therefore, a greedy marginal analysis procedure is used in the literature to solve the service model (see, e.g., Graves [99], Teunter [285]). Such a procedure starts with an empty repair kit and in each iteration one unit of a particular part type is added to the kit until the predefined service level is satisfied. Determining which part type to add is based upon a ratio which measures the relative increase of the service level (i.e., the job fill rate) in relation to the increase of the total holding costs. In previous papers only one unit was added in each iteration. However, when multiple units of the same part type can be used in one job, it is unlikely that adding just one unit is most beneficial in all subsequent iterations. Namely, there are a lot of practical examples in which it is required to replace more than one unit of a part type to fix a job, while replacing only one unit is less likely (see the example at the end of Section 3.1). In such cases it is better to add more than one unit at a time to the repair kit. As a result, a new solution procedure has to be developed to incorporate these possibilities.

The first step of this solution procedure consists of the determination of the order of the number of units to add to the repair kit for each part type. In the second step, a greedy procedure similar to Teunter [285] is used to select the parts that are added to the repair kit based on the increase of the service level. The third, and final, step of the solution procedure consists of improving the solution of step 2 with an improvement and minimization procedure. Each step will be discussed in more detail below. The result of this algorithm is a near-optimal contents of the repair kit (see Section 3.3 for numerical results).

For the first step, we introduce q_k^i as the k-th quantity of part type i to consider in the repair kit, where $q_{k+1}^i > q_k^i$ for all k. The values for q_k^i are set such that the relative increase of the service level (or job fill rate) is decreasing for subsequent values of q_k^i. This is translated into the property formulated in Equation (3.7).

$$\frac{\Delta_i^{job}(q_k^i, q_{k+1}^i)}{q_{k+1}^i - q_k^i} > \frac{\Delta_i^{job}(q_{k+1}^i, q_{k+2}^i)}{q_{k+2}^i - q_{k+1}^i}, \tag{3.7}$$

where $\Delta_i^{job}(q_k^i, q_{k+1}^i)$ represents the increase of the job fill rate when the number of units for part type i increases from q_k^i to q_{k+1}^i. So,

$$\Delta_i^{job}(q_k^i, q_{k+1}^i) = \gamma^{job}([n_1, \ldots, n_i = q_{k+1}^i, \ldots, n_N]) - \gamma^{job}([n_1, \ldots, n_i = q_k^i, \ldots, n_N]).$$

The values of q_k^i for a particular part type i can be found with the following pseudo-code

STEP 1: DETERMINE q_k^i

$n_i = 0$, $q_0^i = 0$, $q_1^i = 1$, $j = 2$, $k = 1$
while $j \le L_i^{max}\overline{M}$
    while $\frac{\Delta_i^{job}(q_{k-1}^i, q_k^i)}{q_k^i - q_{k-1}^i} \le \frac{\Delta_i^{job}(q_k^i, j)}{j - q_k^i}$ and $k > 0$
        $k = k - 1$
    end while
    $q_{k+1}^i = j$, $k = k + 1$, $j = j + 1$
end while

This completes the first step of our solution procedure. The values provided in Table 3.2 show the results for the example of Section 3.1 when applying this step.

i	q_0^i	q_1^i	q_2^i	q_3^i	q_4^i
part 1	0	1	2		
part 2	0	2	4		
part 3	0	1	3	4	6

Table 3.2: The order of the number of units to be put in the repair kit for each part type in the example for the repair kit problem as presented in Table 3.1.

For the second step, we adjust the greedy procedure of Teunter [285] as described at the beginning of this section such that the quantities q_k^i are considered and multiple units of the same part type can be added to the repair kit in one iteration. The pseudo-code for this step is found below.

STEP 2: GREEDY PROCEDURE

1 $n_i = 0$ for all $i \in \{1, \ldots, N\}$, $S = [n_1, \ldots, n_N]$ (empty kit)
2 while $\gamma^{job}(S) < \beta$
3 $\quad i^* = \underset{\{i | n_i < L_i^{max}\overline{M}\}}{\text{argmax}} \left\{ \frac{\Delta_i^{job}(n_i, q_1^i)}{(q_1^i - n_i) H_i} \right\}$
4 $\quad n_{i^*} = q_1^{i^*}$, $S = [n_1, \ldots, n_{i^*}, \ldots, n_N]$
5 $\quad k = 1$
6 $\quad$ while $q_1^{i^*} < L_i^{max}\overline{M}$ and $q_{k+1}^{i^*} < L_i^{max}\overline{M}$
7 $\quad\quad q_k^{i^*} = q_{k+1}^{i^*}$, $k = k + 1$
8 $\quad$ end while
9 $\quad q_k^{i^*} = q_{k+1}^{i^*} = L_i^{max}\overline{M}$
10 end while

Line 1 represents the initialization. In line 2 until line 10 items are added to the repair kit until the required job fill rate is met. In line 3 the part type i^* is selected which adds relatively the most to the repair kit (i.e., it has the highest increase of the job fill rate with respect to the increase of the holding cost). Line 4 adds the units $q_1^{i^*}$ of the selected part type i^* to the repair kit S. In line 5 until line 9 the ordering of $q_k^{i^*}$ is shifted one position, such that $q_1^{i^*}$ represents the next quantity to consider for part type i^*. Table 3.3 illustrates this iterative procedure for the example discussed in Section 3.1.

This greedy procedure immediately stops when the target job fill rate is met. Even though the contents of the repair kit satisfies the service level constraint after performing step 2 of the solution procedure, the total holding costs could

iteration	S	$\gamma^{job}(S)$	$C_H(S)$
0	[0,0,0]	50.40%	0
1	[1,0,0]	55.87%	1
2	[1,2,0]	69.01%	5
3	[1,2,1]	78.30%	11
4	[1,2,3]	95.12%	23
5	[2,2,3]	95.51%	24
6	[2,4,3]	97.00%	28
7	[2,4,4]	98.39%	34
8	[2,4,6]	100%	46

Table 3.3: The results of the greedy procedure for the repair kit problem of Section 3.1.

be reduced when the last iteration is performed in a smarter way. This is the objective of step 3 in the solution procedure. The improvement procedure starts with removing the units which were added to the repair kit S in the last iteration, resulting in repair kit S'. In order to satisfy the service level constraint, items have to be added to S' with the extra constraint $C_H(S') < C_H(S)$ to guarantee a better solution. The same greedy procedure of step 2 can be used to investigate whether a solution S' exists which satisfies the job fill rate criterion with lower holding costs. In the previous pseudo-code, S has to be replaced by S' and line 3 of the pseudo-code should be replaced by

$$3 \qquad i^* = \underset{\left\{ i \,\middle|\, \substack{n_i < L_i^{\max}\overline{M}, \\ C_H(S')+(q_1^i - n_i)H_i < C_H(S)} \right\}}{\operatorname{argmax}} \left\{ \frac{\Delta_i^{job}(n_i, q_1^i)}{(q_1^i - n_i)H_i} \right\}.$$

If such a solution S' exists, it can be investigated for further improvements by setting S to S' and repeating the improvement procedure until no new and better solution is found.

Besides improving S it can also be checked whether units of the current solution S can be removed without replacing them with other parts to reduce the holding costs and still satisfy the job fill rate criterion. This procedure is referred to as the minimization of S. A backtracking procedure is used to check whether the service level is still sufficient when one unit of the last added part type is removed and the one before, and so forth. The repair kit resulting from this minimization procedure is denoted by S''. The overall best solution S^* is found by performing the different procedures in the order shown in Figure 3.1.

When this solution procedure is performed on the same example as described above with a service constraint of 84%, the greedy procedure (step 2) results in

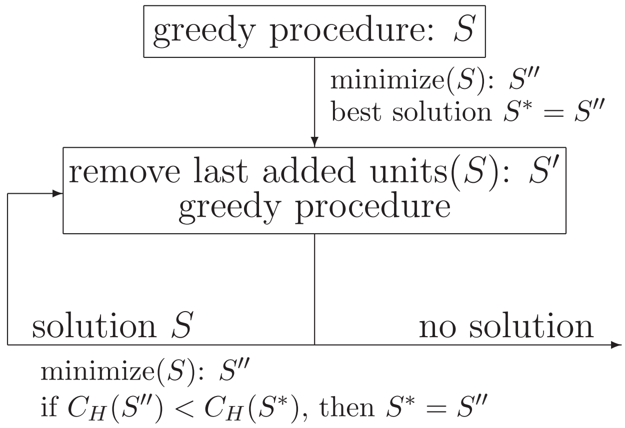


Figure 3.1: The structure of the solution procedure.

$S = [1, 2, 3]$. As illustrated in Table 3.3, the corresponding holding costs for this repair kit are 23 and the job fill rate is 95.12%. The minimization procedure of step 3, transforms this solution in $S'' = [0, 2, 3]$ with holding costs 22 and a job fill rate of 86.36%. This is the best found solution S^* so far. For the improvement procedure, repair kit $S' = [1, 2, 1]$ is used. The same greedy procedure is performed with the extra constraint $C_H(S') < C_H(S) = 23$. This results in $S' = [2, 4, 2]$ as shown in Table 3.4. The minimization procedure reduces this repair kit to $S'' = [1, 4, 2]$ with holding costs 21 and job fill rate 84.16%. This solution corresponds to the optimal contents.

iteration	S	$\gamma^{job}(S)$	$C_H(S)$
3	[1,2,2]	83.03%	17
4	[2,2,2]	83.32%	18
5	[2,4,2]	84.46%	22

Table 3.4: The performance of the improvement procedure in the example for the repair kit problem.

Cost model

The cost model is defined by Equation (3.3). The solution procedure for the service model can be used to solve the cost model in the general setting as explained in Section 3.1. Step 1 of the procedure is similar for both models. However, since the cost model does not have any restrictions upon the service level, the procedure for the cost model needs a different stopping criterion. It also needs to keep track of the solution with the lowest expected total cost. The cost model does not need any improvement steps, since the solution procedure does not stop immediately.

Therefore, step 3 of the algorithm for the service model is removed for the cost model.

Only step 2 of the solution procedure for the service model has to be adapted for the cost model. When the solution with the lowest expected total costs is denoted by S^*, the pseudo-code for step 2 of the solution procedure for the cost model is given below.

Cost model

1 $n_i = 0$ for all $i \in \{1, \ldots, N\}$, $S = [n_1, \ldots, n_N]$ (empty kit), $S^* = S$
2 while $C_H(S) < C_T(S^*)$
3 $\quad i^* = \underset{\{i : n_i < L_i^{max}\overline{M}\}}{\operatorname{argmax}} \frac{\Delta_i^{job}(n_i, q_1^i)}{(q_1^i - n_i)H_i}$
4 $\quad n_{i^*} = q_1^{i^*}$, $S = [n_1, \ldots, n_{i^*}, \ldots, n_N]$
5 $\quad$ if $C_T(S) < C_T(S^*)$ then
6 $\quad\quad S^* = S$
7 $\quad$ end if
8 $\quad k = 1$
9 $\quad$ while $q_1^{i^*} < L_i^{max}\overline{M}$ and $q_{k+1}^{i^*} < L_i^{max}\overline{M}$
10 $\quad\quad q_k^{i^*} = q_{k+1}^{i^*}$, $k = k + 1$
11 $\quad$ end while
12 $\quad q_k^{i^*} = q_{k+1}^{i^*} = L_i^{max}\overline{M}$
13 end while

The new stopping criterion in line 2 is to stop adding items when the holding costs are higher than (or equal to) the expected total costs of the best found solution so far, where $C_T(S^*) = C_H(S^*) + C_{RTF}(S^*)$.

3.3 Numerical results

In this section the performance of the solution procedures as described in Section 3.2 is tested by means of three test cases. Teunter [285] considers two kind of test cases: small instances and large instances. In the definition of small instances at most 8 different part types are used and the maximum tour size is set to 4. For large instances at most 100 different part types are considered and the maximum tour size equals 12. As third case, an additional setting is added which is more representative for reality. In this third setting the number of different part types ranges between 500 and 1,000. The test instances are drawn from (discrete) uniform distributions. Table 3.5 shows the specific distributions that are used for the different parameters to randomly generate 1,000 examples for each test case.

For each test case $P(M = m) > 0$ for $\overline{M} - 2 \leq m \leq \overline{M}$, $\overline{M} - 9 \leq m \leq \overline{M}$ and $\overline{M} - 1 \leq m \leq \overline{M}$, respectively. To make sure that $\sum_m P(M = m) = 1$

	small instances	large instances	representative instances
N	discrete uniform[1,8]	discrete uniform[1,100]	discrete uniform[500,1000]
L_i^{max}	uniform[1,4]	uniform[1,4]	uniform[1,3]
$p_i^{job}(j)$	uniform[0,0.2/L_i^{max}]	uniform[0,0.2/L_i^{max}]	uniform[0,0.0005/L_i^{max}]
H_i	uniform[0,0.35]	uniform[0,0.35]	uniform[0,0.05]
$\overline{M}$	discrete uniform[3,6]	discrete uniform[10,12]	discrete uniform[2,3]
$P(M = m)$	uniform[0,1/3]	uniform[0,1/10]	uniform[0,1/2]
β	uniform[85%,95%]	uniform[85%,95%]	uniform[85%,95%]
P	uniform[0,10]	uniform[0,100]	uniform[40,80]

Table 3.5: The distributions for the parameters used to generate the instances for the different test settings.

the remaining probability mass is put on the middle tour size. Also notice that $p_i^{job}(0) = 1 - \sum_j p_i^{job}(j)$. In this section, the results of all three cases are discussed in more detail for the service model and cost model, respectively.

Service model

In the analysis for the service model, the holding costs for the repair kits obtained by our solution procedure are compared to the repair kits resulting from the procedure of Teunter [285]. Two aspects of the solution procedure are tested: (1) the improvement and minimization procedure (step 3 of our procedure) and (2) the greedy procedure for the exact, closed-form expression to calculate the job fill rate. To test the first aspect, the contents of a repair kit is determined according to Teunter [285] and then our improvement and minimization procedure of Section 3.2 modifies this solution. The relative reduction of the total holding costs for the different instances are shown in the first row of Table 3.6. In the second row the relative reduction of the holding costs is presented when the entire solution procedure of Section 3.2 is used (including the exact formula for the job fill rate) and compared to the outcome of Teunter's [285] procedure. The third row presents the relative deviation of the solution found with our procedure compared to the optimal solution, which is found by enumeration. Optimal solutions can only be found for small instances due to the complexity of the problem.

Based on the results shown in Table 3.6, it can be concluded that the improvement and minimization procedure decreases the holding costs on average by almost 5% for small instances. However, with our closed-form expression for the service level we even find a decrease of the holding costs by 5.8% when the same service constraint is satisfied for the small instances. This corresponds to an average deviation of 0.2% from the optimal solution. For the large instances the improvements are less significant. The results for the representative instances

	small		large		representative	
	average	standard deviation	average	standard deviation	average	standard deviation
approx. JFR + improvements	4.68%	8.43%	0.45%	0.55%	1.05%	1.41%
exact JFR + improvements	5.83%	9.16%	1.30%	0.76%	14.18%	4.30%
deviation from optimality	0.25%	1.27%	-	-	-	-

Table 3.6: Reduction of the total holding costs over 1,000 instances for each test setting, when the solution procedure of Teunter [285] is complemented with our improvement and minimization step and when it is compared to our procedure which incorporates the exact job fill rate (JFR). The final row shows the deviation of the results found with our solution procedure from optimality.

show the most significant cost reductions. The reason that the representative scenario benefits the most from the exact job fill rate expression is because of the different principles behind the two solution procedures. The procedure of Teunter [285] adds units to the repair kit based on the potential of each part type to increase the service level, contrary to our procedure which adds units that immediately contribute (relatively) the most to the repair kit. In the representative scenario, the repair kit only contains at most one unit for most of the part types. The potential for each part type is, however, determined based on the contribution of adding more than one unit of that part type to the repair kit. Consequently, this potential is not always realized and other part types are selected in the next iterations of the greedy procedure. The average number of units per part type in the repair kit is much larger for the small and large instances. Therefore, the potential is a better representation of the actual contribution of the part types in these two scenarios. This is also the reason why the improvement and minimization procedure of step 3 in our solution procedure does not show big improvements for the representative instances.

Table 3.7 shows a number of statistics for the different scenarios. The first two rows show the size of the repair kit and the average number of units per part type in the repair kit. The *size* of a repair kit is defined as the number of units in the repair kit (i.e., $\sum_i n_i$). The results for the representative setting show repair kits with the largest size, but these repair kits also contain the most different part types. Consequently, the repair kits of the representative instances contain on average 0.52 units of each part type. Figure 3.2 also shows this relationship where improvements are more significant when the average number of units per part type (i.e., $\sum_i n_i/N$) is small.

Table 3.7 also presents the frequencies that the exact formula for the job fill rate results in a better solution in comparison to the approximation procedure of Teunter [285]. The best solution is found by the approximate job fill

		small instances	large instances	representative instances
average size of repair kit		9.27	290.26	397.47
average value of n_i		1.81	4.86	0.52
frequencies	approx. JFR + improvements is best	4.0%	0%	0.0%
	exact JFR + improvements is best	19.4%	97.4%	93.6%
	same solution	76.7%	2.6%	6.4%
	exact JFR + improvements is optimal	89.3%	-	-
p-values		<1E-06	<1E-06	<1E-06

Table 3.7: Several statistics about the solutions for the service model in the different test settings.

rate procedure in 4.0% of the small instances, while the exact formulation for the job fill rate finds the best solution in 19.4% of the instances. In the remaining 76.7% of the small instances, both methods result in the same solution. Notice that we included the minimization and improvement procedure in Teunter's [285] algorithm to obtain these results. Otherwise, the approximation procedure of Teunter would never have resulted in a better solution. Table 3.7 also shows that our solution procedure with the exact job fill rate finds the optimal solution in 89.3% of the small instances. For the large and more representative cases the best found solution is almost always found with the exact job fill rate. Therefore, it can be concluded that the solution procedure with the exact expression for the job fill rate, as formulated in Section 3.2, significantly outperforms the procedure with the approximation of Teunter [285]. This can also be concluded when a Wilcoxon test is performed in which the null hypothesis specifies that Teunter's [285] procedure performs better. Based on the p-values[1] shown in Table 3.7 the null hypothesis is rejected such that we can conclude that our solution procedure performs significantly better.

Cost model

The performance for the cost model is tested with the same set of experiments as described above for small, large, and representative instances. Since there is no improvement and minimization procedure in the algorithm, the outcome of our solution procedure is compared to the outcome of the procedure developed by Teunter [285]. Table 3.8 shows the relative savings on the expected total cost for all three test cases. This table also shows that there is hardly any deviation from the optimal solution. Table 3.8 also shows the frequencies how many times

[1] The p-value of a test refers to the probability of wrongly rejecting the null hypothesis if it is in fact true. Small p-values suggest that the null hypothesis is unlikely to be true.

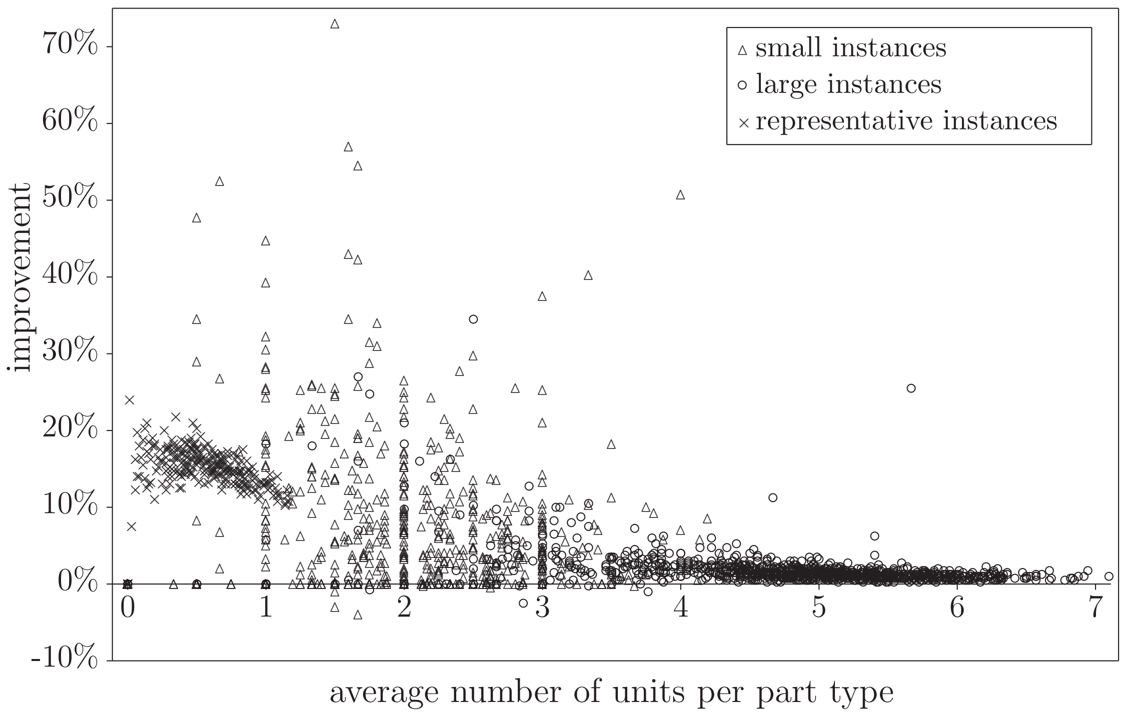


Figure 3.2: The relative improvement plotted against the average number of units in the repair kit per part type for the different test settings.

our solution procedure results in a better solution compared to the procedure of Teunter [285].

3.4 Case study

Besides the test instances of Section 3.3, a case study is performed to get a better feeling for the performance of the solution procedure in practice. In this case study the service model and the cost model is solved and a sensitivity analysis on the service level is performed. This is important since the management of repair service companies wants to know the impact of a particular service level criterion on the holding costs and the size of the repair kit.

In this case study real data from Ricoh Europe have been used. Ricoh is a leading global manufacturer of office automation equipment. They offer products for businesses and for personal use. Ricoh performs the after-sales service to the customers as well. In this case study, we looked at multi-functional systems (combined copier/printer/fax/etc.). Ricoh Europe, located in Amstelveen, is the regional headquarter of Europe, Africa, and the Middle-East. Currently Ricoh Europe has subsidiaries and branches in fourteen countries and factories in France and the United Kingdom. Ricoh Netherlands is one of the subsidiaries. Ricoh Europe has about fifteen thousand distinct types of service parts for about three hundred different multi-functional systems. Ricoh Netherlands has more than 35

		small instances	large instances	representative instances
improvement	average	0.41%	0.50%	2.37%
	standard deviation	1.04%	0.52%	0.85%
deviation from optimality	average	0.00%	-	-
	standard deviation	0.00%	-	-
frequencies	approx. JFR is best	0%	0%	0%
	exact JFR is best	29.7%	87.9%	100%
	same solution	70.3%	12.1%	0%
	exact JFR is optimal	97.8%	-	-
average size of repair kit		10.06	287.20	280.83
average value of n_i		2.30	5.71	0.38

Table 3.8: The results for the cost model.

technicians driving around with a stock value of almost 6,000 Euros each. Ricoh charges RTF cost of 45 Euros if a repair cannot be performed in the first visit.

When the contents of the current repair kits used by the technicians is analyzed, rather low service levels of 53% are observed. Therefore, the expected total return-to-fit costs is quite high. An overview of the current situation is shown in Table 3.9 as well as the results for applying the cost model and the service model and the associated solution procedures.

The solution of the cost model shows an increase of the holding costs by 250%. Despite the fact that the total costs reduce significantly, this solution is undesirable for Ricoh because of a high risk of theft. However, a solution with less holding costs and an improved service level can be found with the service model. Table 3.9 gives an overview on the costs for different values of the service level. Based on these results it is possible to increase the service level by 31% against current holding costs.

In Figure 3.3 the relationship between the total holding costs and the service level is considered. It shows a rapid increase of the service level when the size of the repair kit is small. This concave relationship is what is to be expected based on the property expressed in Equation (3.7).

Figure 3.3 also shows the different costs when units are added to the repair kit. It shows a clear trade-off between the holding costs and RTF costs. Based on the results of this case study it can be concluded that our closed-form expression for the order fill rate and our solution procedure work well in practice. It can help a company decide which parts to put in the repair kit, but it can also help them to analyze their current stock levels.

	job fill rate	holding costs	RTF costs	total costs
current contents	53.14%	1.26	22.23	23.48
cost model	95.99%	3.18	1.91	5.09
service model	84%	1.23	7.59	8.82
	85%	1.30	7.11	8.42
	86%	1.38	6.63	8.01
	87%	1.48	6.09	7.57
	88%	1.56	5.69	7.25
	89%	1.67	5.21	6.88
	90%	1.80	4.72	6.52
	91%	1.94	4.26	6.20
	92%	2.10	3.78	5.88
	93%	2.29	3.31	5.61
	94%	2.53	2.82	5.35
	95%	2.80	2.36	5.16
	96%	3.19	1.90	5.09
	97%	3.75	1.42	5.18
	98%	4.54	0.94	5.49
	99%	5.88	0.47	6.35

Table 3.9: Results for the current contents of the repair kit used in this case study, as well as the results for the cost model and the service model.

3.5 Concluding remarks

Customer-oriented markets become more and more important and, therefore, after-sales services as well. One particular service is a repair service on location, in which a customer is only satisfied when a repair is completed. This means that a technician should have enough spare parts taken along to the customer. If one of the required parts is missing, the technician has to return later and none of the required parts that are available are taken out of the repair kit. As illustrated in Chapter 2, this latter characteristic of the repair kit problem is not dealt with in previous literature. In Chapter 3, we derived an exact, closed-form expression for the service level in a general setting where multiple units of each part type can be used in a job and multiple jobs are performed before the car is restocked. Two procedures are developed to solve the service and cost model which incorporate this exact, closed-form expression for the job fill rate. Based on test instances it can be concluded that this solution procedure performs significantly better com-

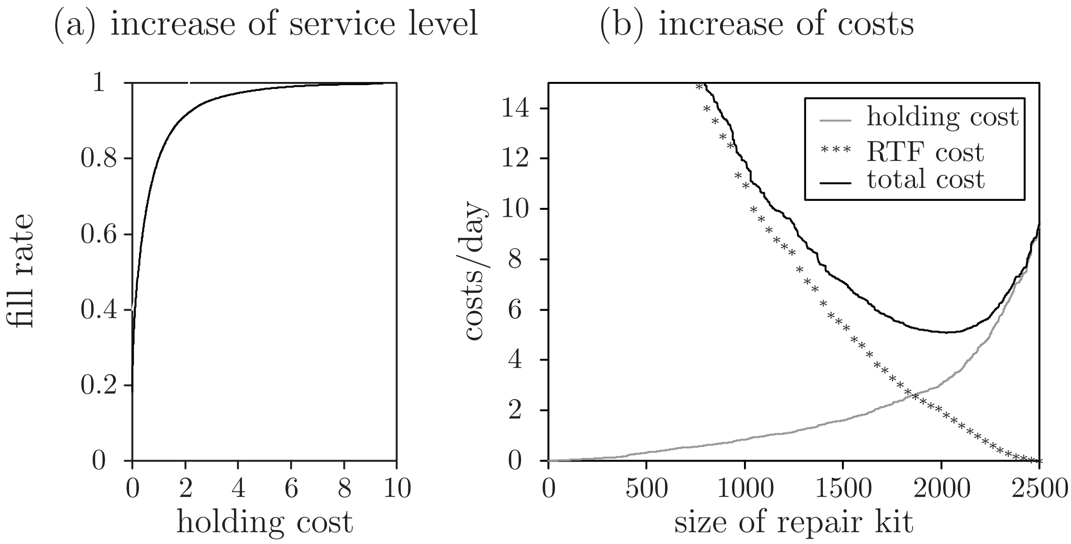


Figure 3.3: The results for the case study: (a) the increase of the service level when the holding costs increase, (b) the different costs when the number of items in the repair kit increases.

pared to other existing procedures in the literature. Especially when only a few units are required in the repair kit for each part type. It finds solutions within a range of 0.2% from the optimal solution. A case study has been used to show the applicability of the solution procedure in practice. Since a cost and service model is considered, we have contributed to the development of customer-focused inventory models for after-sales services. Therefore, this part helps to bridge the gap between theory and practice as discussed in Chapter 1.

Part II

Lost-sales inventory systems

Chapter 4

Introduction to lost-sales inventory systems

As mentioned in Chapter 1, it is important to know the customer's behavior towards stock outs. These behavioral characteristics have to be incorporated in a model such that it represents reality. In most real-world inventory systems excess demand is lost, especially in retail environments where customers can choose from a wide range of items. A customer either buys a substitute product or goes to a different store to buy the product when it is out of stock (as discussed in Section 1.2). Therefore, the original demand is usually lost instead of being backordered. The goal of part II is to model inventory control systems where excess demand is lost such that (near) optimal stock levels can be determined.

In this chapter the available literature on lost-sales inventory control models is presented, and the gap between literature and practice is identified. We start with more background information on lost-sales inventory models in Section 4.1 and compare them with backorder models. This section includes a discussion on why optimal policies are difficult to find for lost-sales models, and how this is dealt with based on observations derived from the literature. Next to that, we propose a methodology that can be used to model and solve lost-sales inventory control problems in general. The available lost-sales inventory models in the literature are classified according to the characteristics of the inventory reordering process. Section 4.2 until Section 4.5 discuss the literature from each of the classes. Section 4.2 provides an overview on the models with continuous reviews, whereas periodic review models are discussed in Section 4.3. First a general description of the developments for both types of replenishment review processes is provided, after which the available literature is discussed in more detail. Models which assume a mixture of lost sales and backordering are addressed in Section 4.4. More related research on lost-sales inventory systems with specific characteristics is provided in Section 4.5. Our conclusions regarding the literature overview and the gap between theory and practice are addressed in Section 4.6. The contribution and outline of this part of the thesis on lost sales are discussed in Section 4.7.

4.1 Lost-sales systems as research area

Inventory models with a backorder assumption have received by far the greatest attention in inventory literature. This is mainly because order-up-to policies are proven to be optimal for backorder models with periodic reviews by Scarf [258]. This optimality result has been studied extensively, and resulted in modifications and extensions (see, e.g., Zabel [321], Veinott [296], Johnson [139]). Different exact algorithms and approximations have been developed to find optimal or near-optimal values for the reorder level and order-up-to level. Some examples are Federgruen and Zipkin [85], and Zheng and Federgruen [326]. A comparison of the different procedures is performed by Porteus [236]. Similar developments are found for replenishment policies with fixed order sizes (e.g., Federgruen and Zheng [84]). Only a few papers include a service level restriction next to a cost minimization objective (e.g., Tijms and Groenevelt [287]). This latter type of backorder models is commonly used in practice. However, the backorder assumption for excess demand is not realistic in many retail environments. When the lost-sales system is approximated by a backorder model, the cost deviations can run up to 30% (see, e.g., Zipkin [328]). Therefore, the customer's behavior has to be modeled with a lost-sales model instead of a backorder model. The objective of this section is to explain the differences between the two types of models, and to show why lost-sales models require a different approach to analyze inventory systems compared to backorder models. We develop a general methodology to model and solve lost-sales inventory systems at the end of this section.

In a backorder model, the inventory position is used as main indicator of the inventory status. It increases when an order is placed and decreases when a demand occurs. Notice that backorders are included in the definition for the inventory position. When the demand is lost instead of backordered, the inventory position does not decrease if the system is out of stock. It is no longer true that the amount of on-hand inventory after the lead time equals the inventory position after the order placement minus the demand during the lead time. Contrary to the backorder model, it is not possible to track the changes in the inventory position independently of the on-hand inventory level when excess demand is lost.

Consequently, a lost-sales model has to keep track of the available inventory on hand at the beginning of a review period, and the quantities of the individual outstanding orders that were placed in past periods and have not yet arrived. As a result, the information vector for a lost-sales model has a length equal to the lead time. Consequently, the state space to describe the inventory system grows exponentially fast with the length of the lead time. Therefore, inventory models with a lost-sales assumption on excess demand are more difficult to analyze compared to models where excess demand is assumed to be backordered. In order to keep the analysis tractable, almost all exact approaches assume that at most one (or two) order(s) can be outstanding at the same time.

To perform an analysis on lost-sales inventory systems, a general methodology can be derived from the papers that model and solve such systems. Such a methodology consists of five phases. Each of these phases is described and clarified below.

1. *identify the characteristics of the inventory replenishment system*: The reordering process is characterized by the review interval (continuous or periodic), the determination of the order size (fixed or variable), and the order costs (with or without fixed order cost). Based on these characteristics of the replenishment process we propose the classification scheme as presented in Table 4.1 to classify the literature. These characteristics have been discussed in more detail in Section 1.1. Furthermore, we distinguish between models with a total cost minimization objective function (*cost model*) or with a service level constraint (*service model*).

2. *identify the assumptions of the inventory model to represent the system*: e.g., the demand distribution, lead time (deterministic or stochastic), the maximum number of outstanding orders. When more than a single order can be outstanding at the same time and lead times are stochastic, difficulties are encountered in properly representing the lead time as a random variable. This is because in practice, orders are almost always received in the same sequence in which they were placed (i.e., orders cannot cross). When the lead time is, however, assumed to be an independent random variable, orders are allowed to cross in time. There is no easy solution to model this problem of dependency between the lead times of the orders outstanding. Therefore, it is common to assume independent lead times when they are stochastic. This assumption makes sense in most real-world applications, where the time interval between placement of two or more orders is usually large enough that there is no interaction between orders. Consequently, it is a good approximation to treat the lead time as an independent random variable while simultaneously assuming that orders do not cross.

3. *develop a Markov model* to represent the on-hand inventory level and the

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">order moment</td></tr>
<tr><td>continuous review</td><td>periodic review</td></tr>
<tr><td rowspan="2">order size</td><td>fixed</td><td>(s, Q)</td><td>(R, s, Q)</td></tr>
<tr><td>variable</td><td>no fixed cost: $(S-1, S)$
fixed cost: (s, S)</td><td>no fixed cost: (R, S)
fixed cost: (R, s, S)</td></tr>
</table>

Table 4.1: The notation for the six types of replenishment policies most often applied in literature and practice as used in this paper to classify lost-sales inventory models.

individual orders outstanding based on the characteristics and assumptions.

4. *analyze the long-run behavior of the inventory model* based on the transition probabilities and steady-state behavior of the model. The stationary distribution function of the on-hand inventory level is used to express the expected average costs and fill rate.

5. *determine the values of the inventory control variables* such as the reorder level and order quantities: Either an exact procedure or an approximation procedure can be used to set these values. Two types of exact procedures are commonly used in literature, namely a policy iteration algorithm or an extensive numerical search procedure. Such procedures can be performed more efficiently if convexity results are derived for the objective function. However, the amount of computational effort remains large. Therefore, approximation procedures are commonly derived. There are also two types of heuristic approaches that are commonly found in the literature. Either the EOQ model of Harris [107] is used to determine the order quantities or the backorder model is used to approximate the steady-state behavior of inventory systems with lost sales. The performance of both heuristic approaches is discussed in Section 4.2 and Section 4.3.

Figure 4.1 summarizes the methodology to model and solve lost-sales inventory control problems. Different models have been developed in the literature to deal with the beforementioned characteristics and difficulties. Assumptions are made about the number of outstanding orders, the demand or lead time distribution. Different replenishment policies also require a different model. This is discussed in Section 4.2 and Section 4.3 for continuous and periodic review systems, respectively.

1. characterize the inventory replenishment system
   - order moment: continuous or period
   - order size: fixed or variable
   - cost structure: no fixed cost or fixed cost
   - objective function: cost model or service model
2. identify the assumptions to develop a model to represent the inventory system
   - demand distribution, lead time, maximum number of outstanding orders
3. develop a Markov model to describe the inventory system
4. analyze the long-run behavior of the inventory model: average cost, service level
5. determine values of inventory control variables
   a. exact approach: policy iteration algorithm, numerical search procedure
   b. heuristic approach: EOQ, backorder

Figure 4.1: The methodology to model and solve lost-sales inventory control problems.

4.2 Continuous review models

The goal of this section is to discuss the available literature on lost-sales inventory systems with continuous reviews. First, the use of the research methodology is illustrated for such inventory systems. Next, characteristics of optimal order quantities are discussed. In the remainder of this section, the available literature on the different replenishment policies are discussed according to our classification scheme of Table 4.1. At the end, a short summary of the developments in the literature is provided as well as some directions for future research on lost-sales systems with continuous reviews.

As mentioned in Section 4.1, it is common to assume that at most one (or two) order(s) can be outstanding at any time when a lost-sales inventory system with continuous reviews is analyzed. Consequently, the Markov model representing the inventory system will be one (or two) dimensional and will consist of the on-hand inventory level (and inventory on order). The decision (or ordering) points in such a model are the time instants at which either a demand occurs and no order is outstanding, or an order is delivered. Next, the transition probabilities and stationary distribution function for the long-run behavior of the on-hand inventory level are derived conditionally on the number of outstanding orders. This distribution function is used to analyze the inventory system in terms of expected cost and service level. The analysis is, however, different for different demand and lead time distributions and it also depends on the replenishment policy. In this section, we discuss different solution techniques available in the literature to find optimal or near-optimal order quantities. In practice, it depends on the specific situation which model to choose. For instance, sales data have to be analyzed to determine the most appropriate demand distribution.

Not much is known about an optimal policy for continuous review models with lost sales. Johansen and Thorstenson [135, 136] are two of the few authors who derive characteristics for optimal order quantities. The authors assume at most one order to be outstanding at the same time in their model and determine the optimal order quantities $a^*(x)$ where x equals the on-hand inventory level at a decision point. They prove the following property for the optimal order quantities:

$$\begin{aligned} &a^*(x+1) \le a^*(x) \le a^*(x+1)+1, && \text{for } x = 0, 1, \ldots, s-1, \\ &a^*(x) = 0, && \text{for } x > s. \end{aligned} \tag{4.1}$$

This means that the optimal order quantity $a^*(x)$ is decreasing in the on-hand inventory level x, and the related rate of decrease is less than one. Furthermore, they show with a numerical example that the optimal policy is prescribed neither by an (s, Q) policy nor by an (s, S) policy. In this section it is shown that most research is performed on such policies. Consequently, there is no comparison of the performances for such policies and an optimal policy in the literature.

In this section, the available lost-sales inventory models with continuous reviews are classified in fixed order size policies and order-up-to policies according to the general classification scheme as used in inventory literature (see Table 4.1). Inventory systems with fixed order size policies are discussed first since the majority of the research is performed on such systems. Next, order-up-to policies are considered where base-stock policies are discussed as a special case. Recall from Section 1.1 that fixed order size policies and order-up-to policies are equivalent to each other in a continuous review setting when the demand size of each individual customer is one (i.e., unit sized). For each replenishment policy we identify the restrictions on the inventory control variables which impose that at most one or two orders can be outstanding at the same time to make the analysis tractable (see also Section 4.1). Besides exact analyses under these assumptions, we also discuss common heuristic procedures to determine the order quantities based on the EOQ formula and the backorder model (see also Section 4.1. The performance of such approximation procedures is compared with exact procedures.

Fixed order size policies

Fixed order size policies are denoted as (s, Q) policies where a new order of size Q is immediately placed when the inventory position drops down to or falls below reorder level s. Let n denote the maximum number of orders that can be outstanding at any time for this policy. The value of n is specified by $(n-1)Q \le s < nQ$ since the inventory position cannot drop below kQ when k orders are outstanding. Consequently, the maximum number of outstanding orders equals the smallest integer strictly larger than s/Q.

The earliest work on lost-sales inventory models with a fixed order size replenishment policy dates back to Hadley and Whitin [105]. They derived an exact expression for the expected total costs under the assumption that there is never more than a single order outstanding (i.e., $s < Q$). This model is extended to deal with stochastic lead times by Ravichandran [242], Buchanan and Love [47], Beckmann and Srinivasan [22], and Johansen and Thorstenson [135] for phase type, Erlang and exponential lead time distributions, respectively. A phase type distribution represents a large class of distributions including the Erlang distribution and a mixture of exponentials. The Erlang distribution corresponds to the exponential distribution when the shape parameter is set to one and to a constant lead time when it is set to infinity. A discounted model for the same inventory system with Erlang distributed lead times is considered by Johansen and Thorstenson [136]. The expected present value of the total inventory costs is minimized instead of the long-run average costs. Numerical results show that discounting is only of importance when the lead time is uncertain (e.g., exponentially distributed), the interest rate is high and the penalty cost for a lost sale is low. A policy iteration algorithm (PIA) is developed by Johansen and Thorstenson [135, 136] to find optimal values of s and Q.

The previous models discussed so far assume a Poisson demand process. More general demand distributions are considered by Kalpakam and Arivarignan [144, 145, 146], and Mohebbi and Posner [202]. In the former models demand is generated from a renewal process, whereas in the latter models demand is not assumed to be unit sized but the demand size of each customer follows an exponential distribution. Optimal values of s and Q can be found with an extensive numerical search procedure. The most general model under the assumption $s < Q$ is developed by Rosling [250]. In this model, any lead time distribution can be used and demand is assumed to be continuous or Poisson distributed. It is only required that the distribution function of the demand during the lead time is log concave. This is the case in most demand distributions suggested for inventory control. An iterative cost minimization procedure is developed to find optimal values of s and Q similar to Hadley and Whitin [105].

The restriction that at most two orders may be outstanding is specified by $Q \leq s < 2Q$. Inventory models with such restrictions are analyzed by Hill [113, 114] for deterministic and Erlang distributed lead times, respectively. Other studies which allow for more than one replenishment order to be outstanding at the same time are Morse [209] for Poisson demand, Kalpakam and Arivarignan [147] for a unit-sized renewal demand process and Mohebbi and Posner [205] for a compound Poisson demand process. These models assume exponentially distributed lead times. Local search techniques are used to find optimal parameter values. Johansen and Thorstenson [137] propose a PIA to find optimal values of s and Q for more general lead time distributions where orders do not cross in time.

Adding a minimal service level restriction to an inventory model with lost sales makes the model more realistic to represent a retail environment, but the

	demand	lead time	assumption	objective
Hadley and Whitin [105]	P	D	$s < Q$	C/S
Ravichandran [242]	P	PH	$s < Q$	-
Beckmann and Srinivasan [22]	P	Ex	$s < Q$	C
Buchanan and Love [47]	P	Er	$s < Q$	C
Johansen and Thorstenson [135, 136]	P	Er	$s < Q$	C
Kalpakam and Arivarignan [144, 146]	uR	Ex	$s < Q$	C
Kalpakam and Arivarignan [145]	M	M	$s < Q$	C
Mohebbi and Posner [202]	CP	Er/HEx	$s < Q$	C/S
Rosling [250]	P/Cont	G	$s < Q$	C
Hill [113]	P	D	$Q \leq s < 2Q$	S
Hill [114]	P	Er	$Q \leq s < 2Q$	S
Morse [209]	P	Ex	-	C
Kalpakam and Arivarignan [147]	uR	Ex	-	C
Mohebbi and Posner [205]	CP	Ex	-	C
Johansen and Thorstenson [137]	P	Er	-	C
Aardal et al. [3]	G	G	-	S

Table 4.2: An overview on lost-sales inventory models with an (s, Q) replenishment policy.

analysis and computations become more difficult. Consequently, hardly any literature is available that studies this problem. Aardal et al. [3] examine a continuous review (s, Q) model with a fill rate constraint. They show by using Lagrange multipliers that any service level restriction implies a penalty cost for lost sales, and they relate the lost-sales model to the backorder model.

An overview of all models with different demand and lead time distributions is provided in Table 4.2. The demand distribution is assumed to be deterministic (D), Poisson (P), compound Poisson (CP), continuous (Cont), Markovian (M) or a unit-sized renewal process (uR). The lead time follows a deterministic (D), exponential (Ex), hyperexponential (HEx), Erlang (Er), phase type (PH), Markovian (M) or general (G) distribution. The fourth column indicates whether at most one or two orders can be outstanding at the same time ($s < Q$ or $Q \leq s < 2Q$, respectively) or no restriction is imposed on the maximum number of outstanding orders. The objective in the models is cost minimization (C) or a service level constraint is included (S).

Based on this overview, we conclude that a lot of research has been performed on (s, Q) replenishment policies for lost-sales inventory systems. However, it remains difficult to analyze the model exactly and determine optimal values of s and Q that minimize the expected total costs C (denoted as s^* and Q^* respectively). In the literature, either a PIA or an extensive search procedure is proposed to find the values of the inventory control variables (see also the methodology of Section 4.1). The computation time can increase rapidly for large inventory systems. Therefore, simple approximation procedures are also developed. Hadley and Whitin [105] derived an approximate expression for the expected total costs in which the expected time period during which the system has no on-hand inventory is negligible. The authors propose an iterative procedure to determine s^* and Q^*, such that the derivative of C in s and Q equals zero (i.e., $\partial C/\partial s = \partial C/\partial Q = 0$). First, they set $Q_1 = Q_w$ where Q_w is the order quantity based on the EOQ formula. The value Q_1 is used to compute s_1 based on $\partial C/\partial s = 0$ and the so obtained value s_1 to compute Q_2 based on $\partial C/\partial Q = 0$, etc. This procedure stops when s and Q are determined with sufficient accuracy. This is called the H-W procedure. Other approximation procedures in the literature are based on the EOQ model where $Q = Q_w$ and s minimizes the expected total costs, or the backorder model is used to determine values of s and Q (see Federgruen and Zheng [84]). In general, the following relationships between these models and the optimal inventory control variables are identified by Hadley and Whitin [105]:

- $Q^* \geq Q_w$, because of the variability in the demand and the holding cost is less than the penalty cost of lost sales.

- $Q^* < Q_{\mathrm{BO}}$ and $s^* > s_{\mathrm{BO}}$, where s_{BO} and Q_{BO} are the optimal reorder level and order quantity in the backorder model. This is because the expected holding costs are lower in the backorder model than in the lost-sales model, since the backorders reduce the average on-hand inventory level (provided that the values of s and Q are the same in both models). But also, the expected penalty costs are higher in the backorder model for the same reason. Consequently, a higher value of s and a lower value of Q are more advantageous in the lost-sales model.

Numerical results show that the cost increases are very minor when one of these heuristic approaches is applied compared to the optimal costs, although the values of the policy variables show larger deviations from using the heuristics (see, e.g., Johansen and Thorstenson [135]). This is due to the flatness of the cost function. The EOQ model results in values of s and Q that are too high and too low, respectively, compared to the exact solutions. The H-W procedure finds too high values of s whereas Q is about the same as in the exact solution. Furthermore, lead time variability shows a significant impact on the values of s^* and Q^* in comparison to the results of the H-W procedure. This is because the H-W procedure ignores the time during which the system is out of stock. With highly variable lead times this is not justified.

Besides fixed order size policies, order-up-to policies are considered in the remainder of this section as included in Table 4.1.

Base-stock $(S-1, S)$ policies

When $Q = 1$ and the demand is unit sized, the (s, Q) policy corresponds to a base-stock $(S-1, S)$ policy with base-stock level $S = s+1$. In this policy every customer's demand is immediately reordered. This policy is optimal in case excess demand is backordered, the lead time is deterministic and no fixed order cost is charged. Karush [157] models a lost-sales inventory systems with this policy as a queueing system. Since the demand process is assumed to be Poisson and the lead times are mutually independent, the author uses the well-known steady-state probability formulas from queueing theory to calculate the average costs per unit time. A convexity property is found in an empirical study by Weinstock and Young [307] when random replenishment lead times of successive orders are correlated. Convexity has been proven differently by Jagers and Van Doorn [128]. An explicit approximation for the base-stock level S is presented by Smith [272]. The use of indifference curves to find an optimal value of S graphically for this inventory system is mentioned by Silver and Smith [269]. These results are applicable for any type of lead time distribution when mutual independency is assumed between lead times. For a compound Poisson demand process, Feeney and Sherbrooke [88] derive the steady-state probabilities for the inventory on order when the $(S-1, S)$ policy is applied. They show that the fill rate in a lost-sales model is higher compared to a backorder model and the inventory on order is always smaller. When the replenishment orders are not able to cross in time, the lead times of simultaneously outstanding orders are dependent. Johansen [133] solves this problem with Erlang distributed lead times. A simple and effective procedure is developed to compute an optimal base-stock level S with respect to the average costs.

Hill [116] shows that the $(S-1, S)$ policy is never optimal if $S \geq 2$. Such policies do not satisfy the optimality properties defined by Equation (4.1). The optimal order quantities are a decreasing function of the inventory level where the rate of decrease is less than one. Base-stock policies have a rate of decrease of exactly one. Therefore, the author proposes a new policy in which a second replenishment order is placed when the lead time remaining on the outstanding order is less than T. Otherwise, the order waits until this occurs. The author calls this the base-stock policy with delay. Consequently, the order process is smoothened over time. This corresponds with the findings of Johansen and Thorstenson [135] on optimal order quantities. If T equals the lead time, then a new order is immediately placed. If T equals zero, at most one order is outstanding at any time. A more detailed policy with delay is discussed by Hill [117], where the value of T depends on the remaining lead time of all outstanding orders. Hence, the value of T has to be determined every time a decision is required (i.e., when a demand is satisfied). A simple to compute lower bound on T is derived. These policies with delay can only be evaluated by simulation. The cost benefits to be gained from delay policies compared to base-stock policies are not large (1-2%) but non-trivial.

Order-up-to (s, S) policies

When a fixed order cost is incurred with each order, a reorder level should be incorporated in the replenishment policy. In case the demand is unit sized, the models for an (s, Q) policy are the same as for an order-up-to (s, S) policy with $Q = S - s$ (as discussed in Section 1.1). To ensure that at most one order can be outstanding in an (s, S) policy, the inventory position should remain above the reorder level s when an order is outstanding, i.e., $S - s > s$. Archibald [13] studies an order-up-to policy with this assumption for a continuous review inventory system with a compound Poisson demand process. An extensive search procedure is required to find optimal values of s and S. An approximate solution is also presented as alternative, where the average satisfied demand during the lead time is added to the reorder level to set the order-up-to level. The exact and approximation procedure are compared with a backorder model and an adjusted backorder model, in which the optimal value of S in the backorder model is decreased by the expected number of shortages during the lead time. The approximation procedure and adjusted backorder model find near-optimal solutions with cost increases less than 0.1%.

In case a service level constraint has to be satisfied, Tijms and Groenevelt [287] provide a simple approximation to set the reorder level for (s, S) replenishment systems with backlogging. The quantity $S - s$ equals the EOQ value. They make a remark on how to modify the model in case excess demand is lost, but never perform any numerical results to validate the model.

An overview of the different demand and lead time distributions for base-stock and order-up-to policies is provided in Table 4.3. The abbreviations are the same as in Table 4.2. Column four indicates whether it concerns a base-stock policy ($s = S - 1$) or an order-up-to (s, S) policy where at most one order can be outstanding at any time ($S - s > s$).

Current state of research on systems with continuous reviews

Based on the overview on continuous review models with lost sales, we conclude that most of the work has been performed on the (s, Q) replenishment policy. The H-W procedure has been extended and improved for different demand and lead time distributions, but it remains difficult to find optimal values of s and Q. Besides exact PIAs and extensive numerical search procedures, several approximation procedures are proposed to find near-optimal reorder levels and order sizes for the cost model. A service model is, however, hardly discussed in the literature.

	demand	lead time	assumption	objective
Karush [157]	P	G	$s = S - 1$	C
Smith [272]	P	G	$s = S - 1$	C
Smith and Silver [272]	P	G	$s = S - 1$	C
Feeney and Sherbrooke [88]	CP	G	$s = S - 1$	C
Johansen [133]	P	Er	$s = S - 1$	C
Hill [116, 116]	P	D	$s = S - 1$	C
Archibald [13]	CP	D	$S - s > s$	C
Tijms and Groenevelt [287]	G	G	$S - s > s$	S

Table 4.3: An overview on the lost-sales inventory models with an order-up-to replenishment policy.

Another topic for future research is the investigation of optimal replenishment policies. There are almost no comparisons with an optimal policy, but only the performance of approximation procedures is illustrated within a specific class of replenishment policies. How well such policies perform compared to an optimal policy is unknown.

4.3 Periodic review models

Optimality results for periodic review inventory models with backorders are well known. As discussed in Section 4.1, this is not the case when excess demand is lost. The majority of the models that include the lost-sales characteristic in a periodic review setting assume no or negligible fixed order cost. Consequently, an order is placed each review period to minimize holding and penalty costs. This results in a regular replenishment process, whereas orders are placed less regular when a fixed order cost is charged. Both situations are observed in practical settings. The classification of models with and without fixed order costs is used in this section as indicated in Table 4.1. This section shows that the main focus of most lost-sales models for periodic reviews is on identifying near-optimal replenishment policies and deriving bounds on optimal order quantities. These results are first derived for the case where the lead time equals the length of a review period ($L = R$), and next for the case where the lead time is an integral multiple of the review period length ($L = nR$). Based on these bounds, myopic replenishment policies are proposed. More common policies to be found in practice and literature are also investigated, as well as some modified policies where a delay in the order process is included. Similar alterations to replenishment policies have already been discussed for continuous review models in Section 4.2. An overview on the

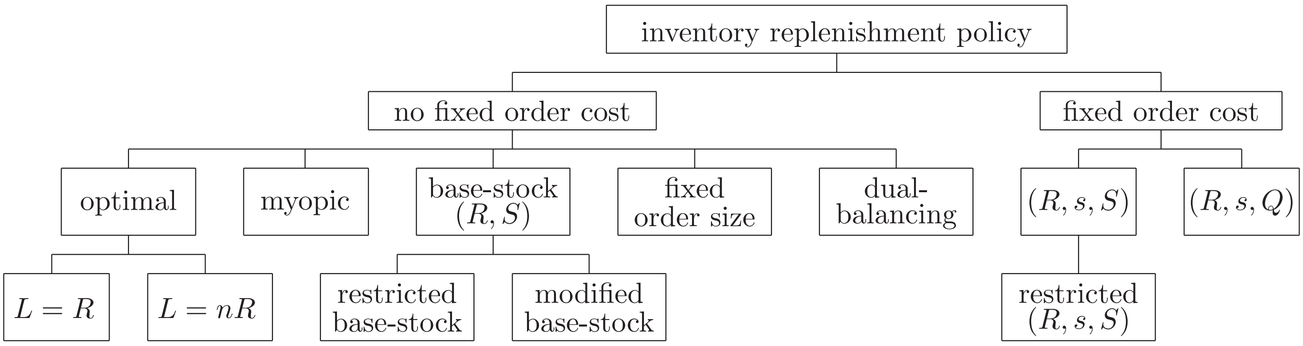


Figure 4.2: The developments in replenishment policies for periodic review lost-sales models.

different types of policies for periodic review lost-sales systems is presented in Figure 4.2. A similar line of research developments is found in the literature when the lead time is stochastic or when the lead time is smaller than or equal to the review period length ($L \leq R$). All literature will be discussed in this section.

No fixed order costs

Bellman et al. [26] are one of the first authors to address the lost-sales inventory control problem in a periodic review system with non-zero lead times. However, they restrict their attention to the special case that the lead time equals one review period. The objective in their model is to minimize the total expected order and penalty costs. Since holding costs are not included, this model is of less relevance in practice. Karlin and Scarf [155] extended this model to the case with holding costs and no restriction imposed on the lead time. They develop optimal dynamic programming equations for the inventory system when excess demand is lost. In their model, the lead time is assumed to be a fixed number of review periods. For the special case that the lead time equals one review period, they show that the cost function and optimal order quantities are well-defined and bounded. Yaspan [318] encounters the same difficulties in deriving properties for optimal order quantities as Karlin and Scarf [155]. The author concludes that an optimal order quantity in a lost-sales system is smaller than in a backorder system, when the same inventory status is observed at a review.

Morton [211] extends the results of Karlin and Scarf [155] to the general case where the lead time is any integral multiple of the review period length. The author derives bounds on the optimal order quantities when linear and proportional holding, penalty and ordering costs are assumed. Notice that this assumption prohibits fixed order costs. The derived upper bound on the optimal order size is used as a myopic policy by Morton [212]. In this policy the order quantity has to be sufficient to fulfill the demand until the delivery of the next order. Denote the probability of not stocking out during this time period by $P_{NS}(x, \mathbf{y}, z)$ where x is the on-hand inventory level, $\mathbf{y}$ the order sizes of the individual orders outstanding and z the order quantity at the current review instant. The value of z is set such that $P_{NS}(x, \mathbf{y}, z)$ equals the fractional benefit of an increase in the order size by one extra unit. If a unit is ordered but not required, the total costs increase with the unit holding cost h. If, on the other hand, the unit is needed, a penalty cost p is saved minus the extra ordering cost c. The myopic policy of Morton [212] prescribes to order z units that satisfies

$$P_{NS}(x, \mathbf{y}, z) = \frac{p-c}{p-c+h}. \tag{4.2}$$

This policy, however, requires to compute the P_{NS} function at each review. A recursive procedure for these calculations is discussed by Morton [211]. If the demand follows a normal distribution, Yaspan [319] shows that multi-dimensional normal tables are required to perform the computations.

Another simple approximation policy is proposed by Morton [212] which resembles the base-stock policy with the exception that the order size cannot exceed the percentile of the demand to be expected in a review period given by the right hand side of Equation (4.2). Let L denote the lead time and D_n the demand during n review periods with cumulative distribution function $F_n(\cdot)$. This approximation policy prescribes to order

$$z = \min\left\{ S - x - \sum_{i=1}^{L-1} y_i, \overline{z} \right\}, \tag{4.3}$$

where

$$S = F_{L+1}^{-1}\left(\frac{p-c}{p-c+h}\right) \quad \Leftrightarrow \quad P(D_{L+1} \leq S) = \frac{p-c}{p-c+h},$$
$$\overline{z} = F_1^{-1}\left(\frac{p-c}{p-c+h}\right) \quad \Leftrightarrow \quad P(D_1 \leq \overline{z}) = \frac{p-c}{p-c+h}.$$

More recently, Zipkin [329] reformulated the original problem formulation of Karlin and Scarf [155] and Morton [211]. The author shows that the optimal cost function is convex, submodular and contains a property related to diagonal-dominance. Such properties are useful to determine optimal order quantities. Additional bounds on the optimal policy are derived, and the model is extended to include limited capacity, correlated demands, stochastic lead times and multiple demand classes. A comparison between optimal replenishment policies in backorder models and lost-sales models is presented by Janakiraman et al. [131]. They show that a lost-sales system has lower optimal costs compared to a backorder system when the cost function is the same in both systems. As a result, a company can offer a discount to induce the customer to backorder without loss of profit when the system has a stock out.

The first research on base-stock policies in a lost-sales setting with periodic reviews dates back to Gaver [92] and Morse [210]. Both authors restrict to cases where the lead time equals the review period length. This model is extended by Pressman [238] to the case where the lead time can be any integral multiple of the review period length. To find an optimal base-stock level, a bisection method can be used due to convexity results on the cost function derived by Downs et al. [74]. They prove that the expected excess demand and on-hand inventory are convex functions of the base-stock level S when fixed lead times are involved. I.e., the cost function is a convex function.

Huh et al. [124] compare the performance of base-stock policies in lost-sales and backorder models similar to the analysis of Janakiraman et al. [131] for optimal replenishment policies. The authors present upper and lower bounds on the total expected costs and optimal base-stock levels in the lost-sales model based on the backorder model with different holding and penalty costs for each unit. They denote the total costs and optimal base-stock level as function of the holding cost h and penalty cost p by $C^S(h,p)$ and $S^*(h,p)$, respectively. A subscript is used to indicate whether a backorder (BO) or lost-sales (LS) system is considered. Then, they prove

$$\begin{aligned} C^S_{BO}(h,p/(L+1)) &\le C^S_{LS}(h,p) \le C^S_{BO}(h,p+Lh) \\ S^*_{BO}(2h(L+1),p-h(L+1)) &\le S^*_{LS}(h,p) \le S^*_{BO}(h,p+Lh). \end{aligned}$$

These bounds on the optimal base-stock level S^*_{LS} are not tight enough to perform well as approximation for the base-stock level. Their main result states that the optimal base-stock level for the backorder model is asymptotically optimal for the lost-sales model,

$$\lim_{p\to\infty} \frac{\min_S C^S_{LS}(h,p)}{C^*_{LS}(h,p)} = \frac{C^{S^*_{BO}(h,p+Lh)}_{LS}(h,p)}{C^*_{LS}(h,p)} = 1,$$

where $C^*_{LS}(h,p)$ is the costs under an optimal policy in the lost-sales model.

Johansen [132] relates the performance of periodic review inventory systems with a base-stock policy to the continuous review model of Smith [272]. The difference with a continuous review model is that orders have to wait in a periodic review system for the next review instant. Therefore, a delay is included in the continuous review model by increasing the demand rate during the lead time. Numerical results illustrate that the fill rate and average on-hand inventory level are approximated very well with this approach compared to the exact values. Besides this base-stock policy, Johansen [132] also derives a policy iteration algorithm (PIA) to find optimal order quantities which minimize the long-run average costs per period. But also, a modified base-stock policy is proposed in which a minimum number of review periods between two subsequent orders is specified. Consequently, the ordering process is smoothened over time. The same idea is found in the delay policy of Hill [116, 117] in case of continuous reviews. A simple heuristic is presented to set the base-stock level and minimal time between orders. This is called the simple modified base-stock policy, and results in near-optimal costs. Unfortunately, the test instances are restricted to values of S less than or equal to four, which is not realistic in many practical settings.

To complete the overview on all replenishment policies, we mention the model developed by Reiman [243] in which a fixed amount is ordered at each review. Such policies do not perform well in practice, since it is a static policy in which the order quantities do not depend on demand occurrences.

In all policies mentioned so far, the sum of the expected holding and penalty costs is minimized. Lost-sales penalty costs incur due to the risk of ordering too few units, whereas holding costs incur due to the risk of ordering too many units. Levi et al. [181] propose a dual-balancing policy, in which the two risks are balanced. The authors prove that the expected total costs of this policy are at most twice the expected costs of an optimal policy. This is the only paper where an upper bound on the cost increase for the expected costs is presented compared to the optimal costs.

Zipkin [328] provides a numerical comparison of several proposed policies from the literature for the cost model. The author concludes that base-stock policies do not perform well, whereas myopic policies are fairly good. Such policies are however not easy to understand or to implement in real life. Therefore, a new type of policy is introduced by Johansen and Thorstenson [138] in which the order size in a pure base-stock policy is restricted to a maximum number,

$$z = \min\left\{ S - x - \sum_{i=1}^{L-1} y_i, q \right\},$$

where q equals $S/(L+1)$ rounded to the nearest integer. They call this the restricted base-stock policy. Notice that such type of policies had already been introduced by Morton (see Equation (4.3)). Numerical results show that the increase in average costs for the best restricted base-stock policy is less than 1% compared to the optimal costs. Since such policies are also simple to implement in real-world applications and they perform well, this type of policies is studied in more detail in Chapter 5.

Stochastic lead times are studied by Nahmias [216] as an extension to the lost-sales model of Morton [212]. To ensure that the lead time is an integral multiple of the review period length, the lead time distribution is reformulated such that orders do not cross as well. Besides the dynamic programming equations to find an optimal policy, an approximation policy is proposed. Janakiraman and Roundy [130] establish some sample-path properties for the lost-sales inventory model with random lead times and a base-stock policy. Similar to Nahmias [216], orders are not allowed to cross in the lead time process. Their main contribution is the convexity of the cost function with respect to base-stock level S. This result justifies the use of common search techniques to determine optimal base-stock levels.

Besides the cost model, Donselaar et al. [289] consider base-stock policies when a prespecified target fill rate β has to be satisfied (i.e., the service model). Demand is assumed to be Erlang distributed and the lead time is fixed. The authors develop a procedure to set the base-stock level. They also propose a dynamic replenishment policy, in which the order size is determined such that the service level constraint is satisfied in the review period after the lead time (i.e., $P_{NS}(x, \mathbf{y}, z) \geq \beta$). Consequently, the order quantities depend on the individual orders outstanding. The amount to order has to be determined at each review instant. This policy can be seen as a myopic policy similar to Morton [212] for the cost model (see Equation (4.2)). Such a dynamic strategy results in a relatively smooth ordering pattern and lower on-hand inventory levels compared to the base-stock policy to obtain a given service level.

In most of the papers mentioned so far, the lead time is assumed to be a fixed or a random integral multiple of the review period. Only a few papers address an inventory system with *fractional lead times*, i.e., the lead time is smaller than the length of a review period. Such a system can be modeled as a one-dimensional Markov chain. Janakiraman and Muckstadt [129] analyze such a model with dynamic programming. They demonstrate that the optimal cost function is convex and they give bounds on the probability of not stocking out while using the optimal order quantity $z^* = a^*(x)$ when x units are on-stock at a review,

$$LB = \frac{p-c-h}{p-c+h} \leq P_{NS}(x, z^*) \leq \frac{p-c}{p-c+h} = UB. \tag{4.4}$$

The optimal policy contains the same property as proven by Johansen and Thorstenson [135, 136] for continuous reviews: $a^*(x+1) \leq a^*(x) \leq a^*(x+1)+1$. Furthermore, the following easily computable bounds are derived

$$\begin{aligned} F_R^{-1}(LB) - x \;&\leq\; a^*(x) \;\leq\; F_R^{-1}(UB), \\ F_{R+L}^{-1}\Big(LB - F_L\big(F_R^{-1}(UB)\big)F_{R-L}(x)\Big) - x \;&\leq\; a^*(x) \;\leq\; F_{R+L}^{-1}(UB) - x. \end{aligned}$$

Numerical results show that the upper bounds are closer to the optimal order quantities than the lower bounds. A myopic policy is proposed that minimizes the expected total costs over the period in which the ordering decision has an immediate impact.

Besides optimal and myopic policies, convexity of the cost function is also proven by Janakiraman and Muckstadt [129] when a base-stock policy is considered in case of fractional lead times. Consequently, a bisection method can be used to determine the best base-stock level. A similar dynamic programming model is formulated by Chiang [56] for the inventory control problem with fractional lead times. The author derives several similar properties of the optimal policy, but the proofs are different. The impact of the review period length on the average on-hand inventory levels and the fill rate is studied through a simulation approach by Sezen [263] for fractional lead times. The numerical results show that the variability in the demand process is the most important factor to set the duration of a review period. No analytical procedure is proposed to determine the length of a review period.

An overview on the different replenishment policies is provided in Table 4.4 for different demand and lead time distributions. The demand is assumed to follow a Poisson (P), Erlang (Er), normal (N) or general (G) distribution. The lead time is either deterministic (D) or stochastic (G). Moreover, the lead time is assumed to be an integral multiple of the review period ($L = nR$) or fractional ($L \leq R$). In the objective function a cost function is minimized (C) or a target service level has to be satisfied (S).

Fixed order costs

When a fixed order cost is incurred with each order in a lost-sales system, the (R, s, S) policy is proven to be optimal when the lead time is zero (see, e.g., Veinott [297, 296], Shreve [265], Bensoussan et al. [28], and Cheng and Sethi [54]). In case of positive lead times, there is no simple optimal replenishment policy. Nahmias [216] is the first author to study such an inventory system and extends the model of Morton [212] to include fixed order costs. In this model the lead time can be deterministic or a random variable. Under the assumption that no more than one order may be outstanding, Hill and Johansen [118] recently proposed a policy iteration algorithm to find optimal order quantities. The authors show that neither an (R, s, S) nor an (R, s, Q) policy is optimal. Both policies are, however, easy to implement in real-world applications. Therefore, some research has been performed on such policies. Notice that the two replenishment policies are not equivalent when the demand is unit sized, contrary to the continuous review systems (as discussed in Section 1.1).

An (R, s, Q) policy is applied in a cost model by Johansen and Hill [134]. The authors assume at most one order to be outstanding at any time. The expected total costs are approximated and the best values of s and Q are determined with a policy iteration algorithm. In the approximation procedure, they introduce a period of risk consisting of the lead time and the undershoot period (time when the inventory position is below the reorder level). The first and second moment of the demand distribution during this risk period are derived, where the demand is assumed to be normally distributed. Consequently, the expected on-hand inventory level and expected demand lost is approximated based on this distribution.

	demand	lead time	assumption	obj	policy
Bellman et al. [26]	G	D	$L = R$	C	optimal
Karlin and Scarf [155]	G	D	$L = R$	C	optimal
Yaspan [318]	G	D	$L = R$	C	optimal
Morton [211]	G	D	$L = nR$	C	optimal
Morton [212]	G	D	$L = nR$	C	myopic, restricted base-stock
Yaspan [319]	N	D	$L = nR$	C	myopic
Zipkin [329]	G	D	$L = nR$	C	optimal
Janakiraman et al. [131]	G	D	$L = nR$	C	optimal
Gaver [92]	G	D	$L = R$	C	base-stock
Morse [210]	G	D	$L = R$	C	base-stock
Pressman [238]	G	D	$L = nR$	C	base-stock
Downs et al. [74]	G	D	$L = nR$	C	base-stock
Huh et al. [124]	G	D	$L = nR$	C	base-stock
Johansen [132]	P	D	$L = nR$	C	optimal, (modified) base-stock
Reiman [243]	G	D	$L = nR$	C	fixed order size
Levi et al. [181]	G	D	$L = nR$	C	dual-balancing
Zipkin [328]	G	D	$L = nR$	C	-
Johansen and Thorstenson [138]	G	D	$L = nR$	C	restricted base-stock
Nahmias [216]	G	G	$L = nR$	C	optimal
Janakiraman and Roundy [130]	G	G	$L = nR$	C	base-stock
Donselaar et al. [289]	Er	D	$L = nR$	S	base-stock, myopic
Janakiraman and Muckstadt [129]	G	G	$L \leq R$	C	optimal, myopic, base-stock
Chiang [56]	G	D	$L \leq R$	C	optimal, base-stock
Sezen [263]	N	D	$L \leq R$	-	base-stock

Table 4.4: An overview on the lost-sales inventory models with no fixed order costs in a periodic review setting.

Tijms and Groenevelt [287] briefly discuss a service model in which a target fill rate has to be satisfied when order-up-to policies are considered in a lost-sales system. The same analysis can be used for periodic review systems as for continuous review systems (see also Section 4.2). However, the numerical results are restricted to a backorder system only.

Fractional lead times are considered by Kapalka et al. [154] where the objective is to minimize the long-run average costs subject to a fill rate constraint. The system is modeled as a one-dimensional Markov chain and the steady-state behavior of the on-hand inventory is studied. A cost objective function is studied by Chiang [57] in case of fractional lead times. No properties of the optimal policy are developed, only an example is provided to illustrate the optimal policy.

Table 4.5 provides an overview on the replenishment policies for periodic review inventory models with lost sales where a fixed cost is incurred with each order. The abbreviations are the same as in Table 4.4.

	demand	lead time	assumption	objective	policy
Nahmias [216]	G	G	$L = nR$	C	optimal
Hill and Johansen [118]	G	D	-	C	optimal
Johansen and Hill [134]	N	D	$L = nR$	C	(R, s, Q)
Tijms and Groenevelt [287]	G	G	$L = nR$	S	(R, s, S)
Kapalka et al. [154]	P	Ex	$L \leq R$	S	(R, s, S)
Chiang [57]	G	G	$L \leq R$	C	optimal

Table 4.5: An overview on the lost-sales inventory models with fixed order costs in a periodic review setting.

Current state of research on systems with periodic reviews

Based on the overview on periodic review models with lost sales, we conclude that most of the developed models focus on inventory systems in which no fixed order cost is charged. In comparison to continuous review models, much more research has been performed on near-optimal order quantities for periodic review inventory systems; upper and lower bounds are derived as well as alternative replenishment policies. Myopic policies perform well for such systems, but they are not insightful for practical use. The restricted base-stock policy of Johansen and Thorstenson [138] is a good alternative replenishment policy. The performance of this policy is unknown for a service model. Furthermore, the lead time is assumed to be an integral multiple of the review period. This is not common in real-life inventory systems. When there are fixed order costs included in a lost-sales model, hardly anything is known about an optimal policy. Therefore, standard replenishment policies are studied. The performance of such policies is not known in comparison to an optimal policy.

4.4 Mixture of lost sales and backorders

Besides inventory systems in which excess demand is either backordered or lost, there are also systems in which a fraction of the excess demand is backordered and the remaining fraction is lost. Such inventory systems encounter similar difficulties to analyze the performance as for lost-sales systems (see Section 4.1). Montgomery et al. [207] are the first authors to analyze inventory systems with this mixture of backorders and lost sales. Similarly, Rosenberg [249] and Leung [180] reformulate this model to simplify the analysis. There is a lot of literature to be found on such partial backorder models. However, most continuous review models focus on (s, Q) replenishment policies whereas most periodic review models focus on (R, S) replenishment policies. In this section we address some recent papers which can be used as point of reference.

Partial backordering models with deterministic demand are studied by Pentico [235, 234] and San José et al. [255]. Models in which the lead-time is considered to be a decision variable are developed by, e.g., Ouyang et al. [227, 228], Liang [183], Chang and Lo [51] and the references therein. In these models, the fraction of backorders is based on a constant backorder probability. This probability can, however, also depend linearly on the number of outstanding backorders (see the overview by Lodree [186] and Hu et al. [122]).

Besides customers that are either willing or not willing to wait for a backorder with some probability, customers can also be willing to wait for some maximum period of time. Posner [237] proposes a model when this time period is stochastic, whereas Das [67] considers a constant patience time. Another possibility to incorporate a mixture of backorders and lost sales in an inventory model is to limit the number of outstanding backorders to a maximum (Krishnamoorthy and Islam [169], Chu et al. [60]). To complete this overview on models with mixtures of backorders and lost-sales, we mention the models in which an incentive is included to backorder a demand during a stock-out period. Such models to prevent lost sales are considered by Netessine [223], Lee [178], and Bhargava [32] and the references therein.

4.5 Related research

From a practical point of view, there are many different inventory systems where excess demand is lost, each with its own characteristics. In the previous sections, we have discussed the inventory literature for single-item systems where demand depletes the inventory levels and orders are delivered after a lead time according to a general replenishment policy. However, there is also literature available on lost-sales inventory systems which include more specific characteristics. Such models are considered in this section.

Unobserved lost sales

When demand is a stochastic variable, it has to be predicted based on historical sales data. A statistical estimation of the mean and variance of the demand should appropriately account for any unobserved demand when sales data is used. Otherwise, the demand is underestimated. Most methods in the literature assume that the demand process follows either a Poisson or normal distribution (see Conrad [64], Sarhan and Greenberg [256, 257], Bell [23, 24], Hill [113], Nahmias [217]). However, Agrawal and Smith [7] show that a negative binomial demand distribution fits their sales data significantly better. The authors propose a parameter estimation methodology that is simple to solve and, therefore, attractive to use in practice. More research on the estimation of demand parameters can be found in Wecker [306], Lau and Lau [175], and Bell [25].

Scarf [259] introduces an empirical Bayesian approach to simultaneously manage inventory control levels and learn about the demand distribution. The author assumes that unmet demand is backlogged. Consequently, the demand is observed. This is obviously not the case for a lost-sales system. Harpaz et al. [106] develop a Bayesian approach for a parameter estimation problem when excess demand is lost. Extensions and improvements are developed by Braden and Freimen [44] and Lariviere and Porteus [174]. A more recent Bayesian approach is developed by Huh et al. [123] and Berk et al. [30]. Other techniques to simultaneously update demand forecasts and inventory control decisions are proposed by Tan and Karabati [283] and Bensoussan et al. [27].

Supply interruptions

As mentioned in Section 4.2, the variability of the lead time greatly influences the optimal values of the inventory control variables, like the reorder level and order quantity. When there are random supply interruptions, the supply process is unreliable and the variability increases. In such inventory systems the source of supply can change from available (or *on*) to unavailable (or *off*). During an unavailable period, the supplier is not able to deliver any orders. One can think of a machine with regular breakdowns. Parlar and Berkin [231] introduce this problem in a continuous review inventory system with lost sales in which the duration of the *on* period is exponentially distributed. They derive the optimal order quantity when the reorder level equals zero for the case of constant and exponentially distributed *off* periods. Their model is further investigated by Bar-lev et al. [19] and Berk and Arreola-Risa [29] when excess demand is lost. The inventory systems considered in the studies mentioned so far assume deterministic demand and zero lead times. An exact analysis of a lost-sales (s, Q) inventory system with unit-sized renewal demands in the presence of supply interruptions is developed by Kalpakam and Sapna [152]. In their model, orders are replenished instantaneously when the supplier is available. Otherwise, the items are supplied at the end of the unavailable period. The *on/off* interruptions in the supply process are assumed to be exponentially distributed. The same inventory system with Poisson demand and constant lead times is studied by Gupta [101]. The author assumes that an accepted order in an *on* period is always delivered regardless of any changes to the supplier's status during the lead time. A more general class of inventory systems with compound Poisson demand and Erlang distributed lead times is investigated by Mohebbi [201].

Emergency replenishments

Besides the regular inventory replenishment process, a supplier can also offer a second means to supply items with a faster mode of resupply at higher costs. This is referred to as *emergency replenishments*. The first continuous review models with lost sales and emergency orders, invoke a replenishment policy where orders of size Q and s are placed when the on-hand inventory level falls down to s and zero, respectively. See for example Morse [209], Bhat [33], Kalpakam and Sapna [149, 150]. A more general type of policies is considered by Mohebbi and Posner [204] where two (s, Q) policies are used for regular and emergency orders where $s < Q$. The numerical results indicate that the emergency mode of resupply is most beneficial when the penalty cost for a stock out is high or the target service level is high.

Moinzadeh and Schmidt [206] study a lost-sales system with emergency replenishments where an $(S-1, S)$ replenishment policy is used. Whether a normal or an emergency order is placed depends on the age of the outstanding orders and the amount of remaining on-hand inventory at the time an order is placed. Kalpakam and Sapna [148] consider a more general order-up-to policy where the inventory position is raised to the order-up-to level S when the on-hand inventory level reaches s or zero. The authors assume no more than one order outstanding of each type at the same time.

Multiple demand classes

In most of the literature on inventory models it is assumed that all demand for a single item is equally important. However, the demand for an item may also be categorized into classes of different importance. For example, demand from key consumers may be given a higher priority than demand arising from less important customers. Another situation where multiple demand classes may be distinguished is a multi-echelon inventory system with emergency orders from lower echelon stocking points. The first lost-sales model with priority demand classes is developed by Cohen et al. [62] for a periodic review system. In each review period, inventory is used to meet the high-priority demand first, and the low-priority demand is satisfied with the remaining inventory. They consider an (R, s, S) replenishment policy and develop an algorithm to set the values of s and S.

It is, however, better to have a critical level policy (introduced by Veinott [295] for a backorder model) where a part of the inventory is kept aside for high-priority demand classes. Such a replenishment policy is considered by Melchiors et al. [199] when excess demand is lost in an inventory system with two demand classes and continuous reviews. The authors adjust the (s, Q) replenishment policy with an extra control variable which prescribes a lower level on the inventory level for which the demand of all demand classes is satisfied. In this policy, low-priority demand is rejected in anticipation of future high-priority demand whenever the inventory level is at or below this prespecified critical level. Isotupa [127] extends the fixed lead time model of Melchiors et al. [199] to the case of exponentially distributed lead times. However, they set the critical level to accept low-priority demand equal to the reorder level. The same model is used by Sivakumar and Arivarignan [271] for a mixture of lost sales and backorders. A continuous review (s, S) policy with a critical level is studied by Lee and Hong [177] where ordered items are resupplied sequentially (i.e., lot-for-lot) according to a 2-phase Coxian distribution. An $(S-1, S)$ replenishment policy with multiple demand classes and lost sales is considered by Ha [103, 104] and Dekker et al. [71]. Consequently, the critical stock level has to be determined for each demand class. Kranenburg and Van Houtum [166] present three very effective heuristic algorithms to solve this problem.

Order splitting

We also mention the work of Hill [115] where a continuous review (s, Q) policy is studied in a lost-sales setting, and an order for Q units is split equally between identical suppliers. The author assumes that at most one order can be outstanding at any time. Order splitting is also considered by Mohebbi and Posner [203] for two non-identical suppliers.

Perishable items

For items with a fixed life time (i.e., perishable items, as introduced in Section 1.1), Schmidt and Nahmias [261] adopt a continuous review $(S-1, S)$ policy for lost-sales models with constant lead times. Perishable inventory systems with exponential life and lead times are analyzed by Kalpakam and Sapna [151]. When excess demand is partially backlogged (see also Section 4.4), we refer to San Jose et al. [253, 254], Yang et al. [316], Abad [4] and the references therein for a recent overview on inventory control models with different deterioration circumstances.

Pricing and inventory control

When the demand depends on the price of an item, the inventory control model should incorporate the selling price as a decision variable. Zabel [322] performs an analysis for such systems in case of lost sales. The author concludes that the lost-sales and backorder assumption share common features. For a recent summary of the developments in pricing and inventory control models with lost sales, we refer to Chen et al. [52] and the references therein. Gavirneni [93] considers a problem where the selling price fluctuates over time whereas the demand distribution is stationary. Pricing and inventory control in a partial backorder model is always considered for perishable items (see the previous subsection on perishable items).

Joint replenishments

When fixed order costs are shared among all items that are ordered simultaneously, there is a cost benefit to place orders jointly. Such multi-item inventory systems are studied by Goyal [96] and Yadavalli et al. [315] in case excess demand is lost.

Multi-echelon

Consider a two-echelon inventory system with one central warehouse and an arbitrary number of retailers. The retailers face customer demand and replenish their stocks from the central warehouse. The warehouse, in turn, replenishes its stock from an outside supplier. The only literature for such inventory systems when excess demand is lost, is restricted to Nahmias and Smith [218], Anupindi and Bassok [10], Andersson and Melchiors [9], Jokar and Seifbarghy [142], Seifbarghy and Jokar [262].

Production

In this overview we do not consider production-inventory control models. The decision issues for such systems are beyond the scope of this thesis. Such decisions concern for instance production rates (Doshi et al. [73], De Kok [70], Nobel [226]), setup times and batching in a multi-product system (Grasman [98], Gurgur [102], Kim and Van Oyzen [160], Krieg and Kuhn [167, 168]), scheduling of maintenance activities (Kenne et al. [158]).

4.6 Lost sales in practical settings

From the literature overview on lost-sales inventory systems, we conclude that not much is known about an optimal replenishment policy when excess demand is lost. The properties and numerical results that have been derived for the optimal order quantities show that there is no structure for an easy-to-understand optimal replenishment policy which can be implemented in real-life applications. Therefore, it is important to first identify the characteristics that are generally observed in practical situations and relate this to the literature, before we present our contribution to the field of lost-sales inventory systems. This gap between theory and practice is the main topic of this section.

Most lost-sales models in the literature focus on continuous review systems. However, in practice, most suppliers dispatch their delivery trucks to retailers at fixed time instances (e.g., once or twice a week). Therefore, it makes sense for retailers to place orders at the same regular basis. Otherwise, the time between ordering and delivery (i.e., the lead time) increases with this extra waiting time at the supplier. Even though the status of the inventory levels can be monitored continuously in time with current information systems at retailers, the actual replenishment process is based on periodic reviews in practice. To our knowledge, all literature on lost-sales inventory models assume lead times to be an integral multiple of the review period (e.g., Zipkin [328], Huh et al. [124]) or fractional lead times (Kapalka et al. [154], Chiang [56, 57]). See also Table 4.4. In practice, the lead time can be of any length. Especially since the lead time should include the time for a retailer to transship the delivered items to the shelves. Therefore, the lead time is not strictly related to the review period length in real inventory systems.

Furthermore, it is common in literature to charge holding costs after demand has occurred in a review period (Zipkin [327]). I.e., demand is assumed to occur periodically at the beginning of a review period. As mentioned by Rosling [251] it is more realistic that demand arrives in continuous time. Consequently, holding costs should be charged over the average on-hand inventory per unit time instead of per review period when the demand has already occurred. Chiang [56, 57] divides the review period in smaller sub-periods and specifies a demand distribution for the sub-periods. In his model, demand occurs periodically at the start of each sub-period. When such sub-periods are small enough, this approach approximates demand occurrences in continuous time but it requires extra computational effort.

To bridge the gap between theory and practice, we focus on periodic review models with lost sales, where the lead time and cost assumptions are relaxed. This is the topic of the next section.

4.7 Contribution and outline of part II

As mentioned in Section 4.6, most real-world inventory systems should be modeled as periodic review systems. The developed models in the literature assume either fractional lead times or lead times to be an integral multiple of the review period length. Another assumption in these models is for the demand and costs to occur periodically at the beginning of a review period. In this thesis we develop general periodic review models in which both assumptions are relaxed. This is referred to as the *new lead time and cost circumstances*. We do not restrict the model to any demand distribution, and only assume fixed lead times which can be of any length. In Chapter 5, we focus on replenishment policies where no fixed order cost is charged for each order. Consequently, pure base-stock policies are commonly used in such situations. As indicated by Johansen and Thorstenson [138], such policies can be extended to a broader class of replenishment policies where an upper limit on the order size is imposed. They call this restricted base-stock policies. Such policies are as simple to implement in practice as pure base-stock policies. Therefore, we consider such policies in more general models with either a cost or service objective. The same analysis is performed in Chapter 6 in case a fixed order cost is incurred with each order. We do not only consider the traditional (R, s, S) and (R, s, Q) policies but we also introduce a new and broader class of replenishment policies. We call such policies restricted order-up-to policies, since the order size is restricted to a maximum. We compare these policies with optimal policies without restrictions on the maximum number of orders outstanding. To our knowledge, we are the first authors to perform such an investigation for large lead times. Besides this comparison between optimal and 'easy-to-understand' policies, it is also important to determine the values of the inventory control parameters for the policy such as the reorder level and order quantities. An exact procedure requires a big computational effort and long computation times (see Section 4.1). Therefore, we also develop approximate expressions for the steady-state probabilities of the on-hand inventory level to analyze the inventory system and find near-optimal values of the inventory control variables. This procedure can be performed for a cost or service model. Consequently, we satisfy all requirements to include lost sales in real-world inventory control systems. In addition, we perform a case study in Chapter 7, which shows that specific characteristics of an inventory system can require a different approach. In this chapter we also show how a single-item approach can be used in a multi-item context with capacity limitations. We decided not to include mixtures of lost sales and backorders in the models developed in this part of the thesis, such systems are studied in part III.

Chapter 5

Lost-sales systems with no order costs

From the literature overview in Chapter 4, it is clear that there is not much understanding of an optimal replenishment policy for lost-sales inventory systems. However, excess demand is lost in many practical settings, especially in a retail environment. Therefore, it is interesting to investigate the influence of lost sales on the performance of inventory systems. In the literature, different approximate policies are proposed as alternative for the optimal policy. Most available approximate policies are not easy to understand or simple to implement in real life. The goal of this chapter is to develop mathematical models for inventory systems with different and more general replenishment policies which are applicable in practice.

As discussed in Section 4.7, most real-world inventory control problems are classified as periodic review systems. All previous work on such systems assume a relation between the lead time and the review period (either an integral multiple or fractional). This is usually not the case in real inventory systems, where the lead time and review period depend on the supplier/retailer relationship. A retailer and supplier have to make agreements on the lead time and review period. They cooperate and share information to make the supply chain more flexible and responsive to customer requests in a cost-efficient manner. Therefore, it is interesting to study the impact of the length of the review period and lead time on the performance measures such as the average total costs and fill rate. This requires two extensions to existing models. First, the lead time has to be modeled explicitly as any number in relation to the review period, contrary to, e.g., Morton [212] or Zipkin [328]. The second modification in our models is to incorporate an objective function which represents the actual costs per unit time instead of per review period. We model demand and costs to incur in continuous time, whereas holding costs are usually charged after the demand that has occurred during a review period in most periodic review models (see Section 4.7). These extensions are referred to as the *new lead time and cost circumstances*.

In this chapter, we focus on inventory systems with negligible fixed order cost for an item. From a practical point of view, this makes sense if no fixed cost is incurred for an order or when it is shared among a large set of items ordered at the same supplier. We develop a model for a periodic review inventory system with fixed order costs in the next chapter. This classification based on fixed order costs to incur corresponds to our classification of inventory literature (see Table 4.1 and Section 4.3). Besides an exact model to determine optimal order quantities, companies prefer to implement easy to understand replenishment policies because of coherency in the replenishment process. Recently, Zipkin [328] performed a comparison of different approximate policies for lost-sales inventory systems where no fixed order cost is incurred. The author concludes that myopic policies perform in general better than pure base-stock policies (PBSPs) with respect to a cost minimization function. However, the latter type of policies is asymptotically optimal when the penalty cost for a lost-sales occurrence is high (see Huh et al. [124]). Such base-stock policies are also fairly simple to apply in real inventory systems, whereas myopic policies are less insightful and they require extra computational effort at each review. See Section 4.3 for more details on both types of policies. Johansen and Thorstenson [138] introduce restricted base-stock policies (RBSPs) as alternative replenishment policy, which imposes an upper limit on the order size in a base-stock policy. A similar type of policy is proposed by Morton [212] (compare Equation (4.2) to Equation (4.3)). RBSPs result in near-optimal order quantities and they are easy to implement in practice. Therefore, our focus is on such policies as well.

In Section 5.1, we introduce the notation and assumptions to model the periodic review inventory control systems with lost sales and fixed lead times under the more general lead time and cost circumstances. The actual inventory models are developed in Section 5.2 for different replenishment policies. We consider the optimal policy, PBSP, and RBSP. Besides a cost model, we also formulate a service model in which a fill rate constraint should be satisfied. Based on a dynamic programming formulation, we use value iteration to compute the performances of interest, such as the fraction of demand lost, the average inventory on hand and the average costs. Even though these models give exact results, the computational effort can be quite excessive, especially for relatively long lead times (see Section 4.1). In Section 5.3 we derive closed-form expressions to approximate the performance of inventory systems with a PBSP. Such approximations can be used to determine near-optimal values of base-stock level S for the cost and service model. We also indicate how this approximation technique can be used to set the maximum order quantity in the RBSP. In Section 5.4 we compare the performance of both base-stock policies to the optimal costs. Section 5.5 contains our concluding remarks.

5.1 Notation and assumptions

In order to model a periodic review inventory system we first introduce the general time framework and demand notations. The inventory system is modeled as a Markov chain, and the performance is analyzed with a dynamic programming formulation. The goal of this section is to present the general framework, whereas the specific replenishment policies are discussed in more detail in Section 5.2.

The time between two reviews is called a review period. Its length is denoted by R. At each review instant an order is issued. The order arrives after a constant lead time L. Let T denote the start of a considered review period, and t the time of the order delivery within the same period. Define $r = L \bmod R$. Note that $t = T + r$. The number of full review periods from time T until the delivery of the order issued at time T equals $l = (L - r)/R$. I.e., l equals L/R rounded down to the nearest integer. Hence, $L = lR + r$. Figure 5.1 illustrates the notation based on an example where $R \leq L < 2R$ (i.e., $l = 1$). The special case when the lead time is an integer multiple of the review period is described by Zipkin [328] and Johansen and Thorstenson [138]. In their papers $r = 0$, $l = L/R$, and $t = T$.

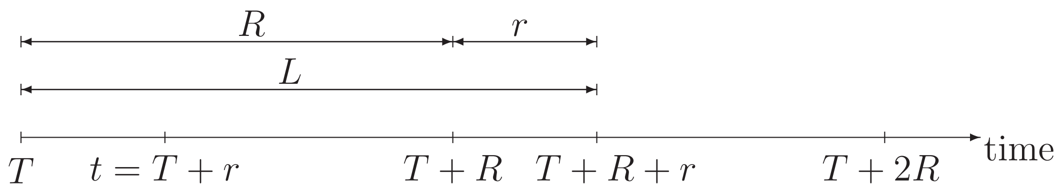


Figure 5.1: An example of the time framework for reviews and order deliveries where $R \leq L < 2R$.

Demand is assumed to be independent and identically distributed over time. We consider two types of discrete demand distributions: compound Poisson and negative binomial. In the first distribution, customers arrive according to a Poisson process with rate λ. When the demand size is unit sized, the demand process over a certain period of time is called a pure Poisson process. Moreover, when the number of units demanded by each customer is a stochastic random variable, the demand process is called a compound Poisson process. In practice it is common to observe that the customer's demand size follows a delayed geometric distribution with mean $\mu = 1/(1 - \theta)$, where $\mu \geq 1$ (Johnston et al. [141]). This specific type

distribution	$E[D]$	$Var[D]$
compound Poisson	$\lambda\mu$	$\lambda(2\mu^2-\mu)$
negative binomial	$w\frac{1-u}{u}$	$w\frac{1-u}{u^2}$

Table 5.1: The mean $E[D]$ and variance $Var[D]$ of the demand per unit time for the compound Poisson and negative binomial demand distribution.

of a compound Poisson process is referred to as a stuttering Poisson process in the literature. The total demand of n customers is denoted as X_n, where

$$P(X_1 = d) = (1-\theta)\theta^{d-1}, \qquad d \geq 1, \tag{5.1}$$

and more generally

$$P(X_n = d) = \begin{cases} 1-\theta, & n = 1, d = 1, \\ \theta P(X_n = d-1), & n = 1, d > 1, \\ \theta P(X_n = d-1) + (1-\theta)P(X_{n-1} = d-1), & n = 2, \ldots, d-1, \\ (1-\theta)P(X_{n-1} = d-1), & n = d, \\ 0, & n > d. \end{cases}$$

Furthermore, the demand during a time period of length τ is a stochastic random variable D_τ with probability mass function $g_\tau(d)$, where

$$g_\tau(d) = \begin{cases} e^{-\lambda\tau}, & \text{if } d = 0, \\ \sum_{n=1}^{d} P(X_n = d)e^{-\lambda\tau}\frac{(\lambda\tau)^n}{n!}, & \text{if } d > 0. \end{cases}$$

In the special case of a pure Poisson demand process, $\mu = 1$ and $P(X_n = n) = 1$.

In the second type of demand distribution, demand follows a negative binomial distribution, where

$$g_\tau(d) = \binom{d+w-1}{d} u^w (1-u)^d.$$

Such demand distributions are also common to be observed in practice (see Agrawal and Smith [7]). The mean and variance of the demand per unit time is denoted as $E[D]$ and $Var[D]$, respectively. They are presented in Table 5.1 for both demand distributions.

We also define,

$$\begin{aligned} \mathcal{G}^0_\tau(i) &= Pr(D_\tau < i) = \sum_{d=0}^{i-1} g_\tau(d), \\ \mathcal{G}^1_\tau(i) &= E\big[(i-D_\tau)^+\big] = \sum_{d=1}^{i} \mathcal{G}^0_\tau(d), \end{aligned}$$

with $(A)^+ = \max\{A, 0\}$.

We denote a set of non-negative integers by $\mathbb{N}_0$, and a set of all integers between m and n by $\mathbb{N}_{m,n} = \{i \in \mathbb{N}_0 \,|m \leq i \leq n\} = \{m, m+1, \ldots, n\}$. Furthermore, $\mathbb{N}^{l+1}_{m,n}$ is defined as the $(l+1)$-fold cartesian product of $\mathbb{N}_{m,n}$. This notation is used to model the inventory system as a Markov chain. Its state at time T (just before ordering) is denoted $(i, \mathbf{y})$, where i is the actual inventory on hand and $\mathbf{y}$ is a vector with components $y_k,\ k \in \mathbb{N}_{0,l-1}$. Component y_k is the number ordered at time $T-(l-k)R$ to be delivered at time $T-(l-k)R+L = t+kR$ (see also Figure 5.1). We let $\mathbf{F}(\mathbf{y})$ denote the vector obtained from $\mathbf{y}$ by removing its first component. The state of the Markov chain at time t is denoted $(j, \mathbf{z})$, where j is the updated on-hand inventory level and $\mathbf{z} = (\mathbf{F}(\mathbf{y}), y_l)$. The component y_l represents the amount ordered at time T. The replenishment policy prescribes how y_l depends on $(i, y_0, \ldots, y_{l-1})$. We denote the actual demand from time T to time t and from t to $T+R$ by d_r and d_{R-r}, respectively. Hence, $j = (i-d_r)^+ + y_0$. Consequently, the state space is an $(l+1)$-dimensional vector where the first component specifies the on-hand inventory level and the remaining l components represent the orders outstanding.

To complete the Markov chain description, we have to specify the one-step transition probabilities between the different states of the inventory system. The transition probabilities from state $(i, \mathbf{y})$ at time T to state $(j, \mathbf{F}(\mathbf{y}), y_l)$ at time t is denoted by $P_{(i,\mathbf{y}),(j,\mathbf{F}(\mathbf{y}),y_l)}$. These transition probabilities depend on the replenishment policy. Similarly, we denote the transition probabilities from state $(j, \mathbf{z})$ at time t to state $(i, \mathbf{z})$ at time $T+R$ by $p_{(j,\mathbf{z}),(i,\mathbf{z})}$. During this period of length $R-r$, the inventory level only decreases because of demand occurrences and no action is taking place regarding any order. Therefore, it is easy to see that the one-step transition probabilities at time t are

$$p_{(j,\mathbf{z}),(j-d,\mathbf{z})} = \begin{cases} g_{R-r}(d), & \text{if } 0 \leq d < j, \\ 1 - \mathcal{G}^0_{R-r}(j), & \text{if } d = j, \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

To compute the total expected costs over a review period, we express the expected costs incurred over a period of length τ by $c_\tau(i)$ when the on-hand inventory level equals i at the beginning of this period and no order delivery occurs during this period. The cost function $c_\tau(i)$ consists of the expected holding and penalty costs. First, let $H(i)$ be the mean time-weighted stockholding until the inventory is fully depleted given an initial inventory level of i units. By definition $H(0) = 0$. For the compound Poisson demand process where the customer demands follow a delayed geometric distribution,

$$H(i) = \frac{i}{\lambda} + \sum_{j=1}^{i-1} P(X_1 = j) H(i-j) = \frac{i}{\lambda} - \theta \frac{i-1}{\lambda} + H(i-1), \tag{5.3}$$

because it takes on average $1/\lambda$ time units before a customer arrives and $(i - X_1)^+$ units remain on stock after this demand arrival. The mean time-weighted inventory during a period of length τ for a given initial inventory level of i units is

$$H_\tau(i) = H(i) - \sum_{j=0}^{i-1} g_\tau(j) H(i-j). \tag{5.4}$$

Equation (5.3) is only true for a Poisson customer arrival process, since it is based on the PASTA property (see Wolff [308]). Otherwise, the average on-hand inventory level per unit time is approximated by the average of the start and end inventory for this period of length τ. So,

$$H_\tau(i) \approx \tau \frac{i + E\big[(i - D_\tau)^+\big]}{2} = \tau \frac{i + \mathcal{G}_\tau^1(i)}{2}. \tag{5.5}$$

The expected demand lost during this period is

$$E\big[(D_\tau - i)^+\big] = E\left[D_\tau\right] - i + E\big[(i - D_\tau)^+\big] = \tau E[D] - i + \mathcal{G}_\tau^1(i). \tag{5.6}$$

Let h denote the unit holding cost per unit time and p the unit penalty cost for each lost demand. Consequently, $c_\tau(i) = hH_\tau(i) + pE\big[(D_\tau - i)^+\big]$.

The performance measures of interest for this inventory system are computed by value iteration. Let $V_n(i, \mathbf{y})$ denote the total expected costs incurred over the time interval from time T to time $T + nR$ when the system is in state $(i, \mathbf{y})$ at time T and the system incurs no costs at and after time $T + nR$. Moreover, let $v_n(j, \mathbf{z})$ denote the total expected costs incurred over the time interval from time t to time $T + nR$ when the system is in state $(j, \mathbf{z})$ at time t and the system incurs no costs at and after time $T + nR$. Consequently, $V_0(i, \mathbf{y}) = 0$ and, recursively for $n = 1, 2, \ldots$

$$\begin{aligned} v_n(j, \mathbf{z}) &= c_{R-r}(j) + \sum_i p_{(j,\mathbf{z}),(i,\mathbf{z})} V_{n-1}(i, \mathbf{z}), \\ &= c_{R-r}(j) + \sum_{d=0}^{j-1} g_{R-r}(d) V_{n-1}(j-d, \mathbf{z}) \\ &\qquad + \left(1 - \mathcal{G}_{R-r}^0(j)\right) V_{n-1}(0, \mathbf{z}), \qquad (5.7) \\ V_n(i, \mathbf{y}) &= c_r(i) + \sum_{j, y_l} P_{(i,\mathbf{y}),(j,\mathbf{F}(\mathbf{y}),y_l)} v_n(j, \mathbf{F}(\mathbf{y}), y_l). \qquad (5.8) \end{aligned}$$

The long-run expected costs for this dynamic programming formulation can be computed by value iteration (see, e.g., Tijms [286]). A value iteration algorithm with accuracy number ε repeats to increase n by one and compute the value functions of Equation (5.7)-(5.8) until $M_n - m_n < \varepsilon m_n$, where

$$
\begin{aligned}
m_n &= \min_{(i,\mathbf{y})}\{V_n(i,\mathbf{y}) - V_{n-1}(i,\mathbf{y})\}, \\
M_n &= \max_{(i,\mathbf{y})}\{V_n(i,\mathbf{y}) - V_{n-1}(i,\mathbf{y})\}.
\end{aligned}
$$

When this value iteration algorithm is stopped after the n^{th} iteration, then $\frac{m_n+M_n}{2R}$ cannot deviate more than $100\varepsilon\%$ from the long-run average costs per unit time. This result is denoted as g, and represents the average expected total costs. Notice, when $h = 1$ and $p = 0$ this expresses the average on-hand inventory level, and when $h = 0$ and $p = 1$ this expresses the average demand lost. This latter scenario is used to compute the fill rate $\beta = 1 - g/E[D_R]$. The numerical results reported in Section 5.4 are computed with $\varepsilon = 1\text{E-}4$.

5.2 Different replenishment policies

As mentioned in Section 5.1, the transition probabilities in the Markov chain depend on the replenishment policy. In this section, we develop models for pure and restricted base-stock policies (abbreviated as PBSP and RBSP, respectively), since they are easy to understand and implement in practice. We also show in Section 5.4 that such policies result in near-optimal order quantities. We start with the development of a model where optimal quantities are ordered at each review. Only minor modifications are necessary to model the PBSP and RBSP.

5.2.1 Optimal policy

First, a cost model is developed to find the optimal policy that minimizes the long-run expected total costs. Next, we discuss how the cost model can be used to solve the service model, in which a minimal service level has to be achieved next to a cost objective. The analogy between the cost function and the fill rate has already been mentioned briefly in Section 5.1.

Cost model

To find an optimal policy and the affiliated costs for the cost model, we model the system as a Markov chain with an infinite state space. More specifically, the state space at time T equals $\mathcal{S}^T = \{(i,\mathbf{y}) \in \mathbb{N}_0^{l+1}\}$ and at time t it is $\mathcal{S}^t = \{(j,\mathbf{z}) \in \mathbb{N}_0^{l+1}\}$. At time T, the one-step transition probabilities depend on the entire state description. Therefore, we use the following relative value function at time T

$$\begin{aligned} V_n(i,\mathbf{y}) &= c_r(i) + \min_{y_l \geq 0} \Bigg\{ \sum_{d=0}^{i-1} g_r(d) v_n\big(i-d+y_0, \mathbf{F}(\mathbf{y}), y_l\big) \\ &\qquad + \big(1-\mathcal{G}_r^0(j)\big)\, v_n\big(y_0, \mathbf{F}(\mathbf{z}), y_l\big) \Bigg\}. \end{aligned} \tag{5.9}$$

The value function $v_n(j,\mathbf{z})$ at time t is specified by Equation (5.7).

In the following example demand is assumed to follow a pure Poisson process with $\lambda = 5$ and $\mu = 1$. Furthermore, $R = 1$, $L = 1.5$, $h = 1$, and $p = 19$. Let $y_l = a^*(i,\mathbf{y})$ denote the optimal order quantity when the on-hand inventory level equals i units and the outstanding order quantities are $\mathbf{y}$ at review time T. Notice that in this example at most one order is outstanding at a review. Hence, $\mathbf{y} = \{y_0\}$ in this example. An optimal policy can be found with a value iteration algorithm, and is provided in Table 5.2. The first column represents the on-hand inventory level i, whereas the first row represents the size of the outstanding order y_0 at a review. This optimal policy results in the minimal costs $C^* = 9.63$ and corresponds to a fill rate of 98.05%.

Service model

The service model is more difficult to analyze, since it is not a standard problem. It can be modeled as a constrained dynamic programming (CDP) problem. The solution approach for such a problem is to transform the CDP problem into an unconstrained dynamic programming problem which can be solved using well-known techniques such as the value iteration algorithm as described in Section 5.1. To solve the CDP problem, the method of Lagrange multipliers is employed. Let $\gamma \in [0,\infty)$ denote the Lagrange multiplier. The idea behind this approach is to include the service constraint in the objective function, such that the constrained problem is transformed in an unconstrained problem. This is called the Lagrange relaxation.

$$\begin{array}{ll} \text{minimize} & \text{costs} \\ \text{subject to} & \text{service level} \geq \overline{\beta} \end{array} \quad \Rightarrow \quad \text{minimize} \quad \text{costs} + \gamma(\overline{\beta} - \text{service level})$$

For more information on the connection between the cost model and the service model, we refer to Van Houtum and Zijm [293].

In order to solve the new unconstrained problem, we can use the cost model. Let γ represent the penalty cost p for lost demand in the cost model. Moreover, let $a^*_\gamma(i,\mathbf{y})$ denote the optimal policy in the cost model for a certain value of γ. This policy can be found with a value iteration method as described before. Next, the corresponding fill rate for this policy has to be computed. Therefore, we set

i \ y_0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	8	8	8	8	8	8	8	7	6	6	5	4	3	2	1
1	8	8	8	8	8	8	8	7	6	6	5	4	3	2	1
2	8	8	8	8	8	8	7	7	6	5	5	4	3	2	1
3	8	8	8	8	8	8	7	7	6	5	4	3	2	1	
4	8	8	8	8	8	7	7	6	5	5	4	3	2	1	
5	8	8	8	8	7	7	6	6	5	4	3	2	1		
6	8	8	8	7	7	6	6	5	4	3	2	1			
7	8	8	7	7	6	6	5	4	3	2	1				
8	8	7	7	6	6	5	4	3	2	1					
9	7	7	6	6	5	4	3	2	1						
10	7	6	6	5	4	3	2	1							
11	6	6	5	4	3	2	1								
12	6	5	4	3	2	1									
13	5	4	3	2	1										
14	4	3	2	1											
15	3	2	1												
16	2	1													
17	1														

Table 5.2: The optimal order quantities $a^*(i, y_0)$ when the on-hand inventory level equals i units and the outstanding order has size y_0 at a review instant.

$h = 0$ and $p = 1$ in the cost function such that the result of the value iteration algorithm (denoted by g) represents the average demand lost during a review period. Consequently, the fill rate equals $\beta = 1 - g/E[D_R]$. To compute the value of g, the policy $a^*_\gamma(i, \mathbf{y})$ is used in Equation (5.9) to set the value of y_l in the unconstrained dynamic programming formulation. The value of γ can be determined with a bisection method such that the fill rate equals the service constraint $\overline{\beta}$. Such a procedure is possible because lost demand is penalized more when the Lagrange multiplier γ increases, which results in a higher service level. Therefore, if for a certain value of γ the service level constraint is satisfied ($\beta \geq \overline{\beta}$), this value should decrease and, otherwise, it should increase. This is repeated until the fill rate is sufficiently close to $\overline{\beta}$. This procedure to solve the service model is summarized below. Line 6 prescribes to increase the value of γ by 10 if the service level is not satisfied and the upper bound on γ is not specified. Otherwise, half of the interval on γ is excluded by checking the middle value of the interval on γ.

SERVICE MODEL

set $\gamma = 1$, $\gamma_{\text{LHS}} = 0$, $\gamma_{\text{RHS}} = \infty$
do
  solve the cost model with $p = \gamma$ to find $a^*_\gamma(i, \mathbf{y})$
  compute the expected fill rate β when $y_l = a^*_\gamma(i, \mathbf{y})$
  if $\beta \leq \overline{\beta}$, then
    $\gamma_{\text{LHS}} = \gamma$ and $\gamma = \frac{1}{2}\big(\gamma_{\text{LHS}} + \min\{\gamma_{\text{RHS}}, \gamma + 20\}\big)$
  else
    $\gamma_{\text{RHS}} = \gamma$ and $\gamma = \frac{1}{2}(\gamma_{\text{LHS}} + \gamma_{\text{RHS}})$
  end if
while $(|\beta - \overline{\beta}| > \varepsilon)$

5.2.2 Pure base-stock policy (PBSP)

The PBSP with base-stock level S prescribes to issue a replenishment order at each review time T, such that the inventory position (inventory on hand plus inventory on order) equals base-stock level S. Hence, $y_l = S - i - \sum_{k=0}^{l-1} y_k$ and the state space of the Markov chain at time T is

$$\mathcal{S}_S^T = \left\{ (i, \mathbf{y}) \in \mathbb{N}_{0,S}^{l+1} \,\middle|\, i + \sum_{k=0}^{l-1} y_k \leq S \right\}. \tag{5.10}$$

The state space at time t is

$$\mathcal{S}_S^t = \left\{ (j, \mathbf{z}) \in \mathbb{N}_{0,S}^{l+1} \,\middle|\, j + \sum_{k=1}^{l} z_k \leq S \right\}. \tag{5.11}$$

The one-step transition probabilities from state $(i, \mathbf{y}) \in \mathcal{S}_s^T$ at time T are

$$P_{(i,\mathbf{y}),(i-d+y_0,\mathbf{F}(\mathbf{y}),S-i-\sum_{k=0}^{l-1} y_k)} = \begin{cases} g_r(d), & \text{if } 0 \leq d < i, \\ 1 - \mathcal{G}_r^0(i), & \text{if } d = i, \\ 0, & \text{otherwise.} \end{cases} \tag{5.12}$$

The one-step transition probabilities from state $(j, \mathbf{z}) \in \mathcal{S}_s^t$ at time t are given by Equation (5.2). According to Equation (5.8) we have for all $(i, \mathbf{y}) \in \mathcal{S}_s^T$

$$\begin{aligned} V_n(i, \mathbf{y}) \;&=\; c_r(i) + \sum_{d=0}^{i-1} g_r(d) v_n\Big(i - d + y_0, \mathbf{F}(\mathbf{y}), S - i - \sum_{k=0}^{l-1} y_k\Big) \\ &\quad + \big(1 - \mathcal{G}_r^0(i)\big)\, v_n\Big(y_0, \mathbf{F}(\mathbf{z}), S - i - \sum_{k=0}^{l-1} y_k\Big), \end{aligned} \tag{5.13}$$

whereas Equation (5.7) specifies $v_n(j, \mathbf{z})$ for all $(j, \mathbf{z}) \in \mathcal{S}_s^t$.

To solve the cost model when a PBSP is applied, the value of base-stock level S has to be found which minimizes the expected total costs. Let us denote this value by $\overline{S}$. Numerical results indicate that the cost function is convex. Consequently, $\overline{S}$ can be found with a bisection method. For a service model, $\overline{S}$ equals the smallest value of base-stock level S such that the service level constraint $\overline{\beta}$ is satisfied, since the fill rate is an increasing function in S.

5.2.3 Restricted base-stock policy (RBSP)

The RBSP with base-stock level S and upper limit q on the order size prescribes to issue a replenishment order at each review time T of size

$$y_l = \min\left\{ S - i - \sum_{k=0}^{l-1} y_k, q \right\}.$$

The state space of the Markov chain at time T is $\mathcal{S}_{S,q}^T = \bigcup_{i=0}^S \left(\{i\} \times \mathbb{Y}_{S,q}(i)\right)$, where

$$\mathbb{Y}_{S,q}(i) = \left\{ \mathbf{y} \in \mathbb{N}_{0,q}^l \,\middle|\, \sum_{k=0}^{l-1} y_k \leq S - i \right\},$$

whereas the state space at time t equals $\mathcal{S}_{S,q}^t = \bigcup_{j=0}^S \left(\{j\} \times \mathbb{Z}_{S,q}(j)\right)$, where

$$\mathbb{Z}_{S,q}(j) = \left\{ \mathbf{z} \in \mathbb{N}_{0,q}^l \,\middle|\, \sum_{k=1}^{l} z_k \leq S - j \right\}.$$

Due to the restriction on order size y_l in the RBSP, the one-step transition probabilities from state $(i, \mathbf{y}) \in \mathcal{S}_{S,q}^T$ are

$$P_{(i,\mathbf{y}),(i-d+y_0,\mathbf{F}(\mathbf{z}),\min\{S-i-\sum_{k=0}^{l-1} y_k,q\})} = \begin{cases} g_r(d), & \text{if } 0 \leq d < i, \\ 1 - \mathcal{G}_r^0(i), & \text{if } d = i, \\ 0, & \text{otherwise.} \end{cases}$$

For this policy, the value function at time T equals

$$\begin{aligned} V_n(i, \mathbf{y}) &= c_r(i) + \sum_{d=0}^{i-1} g_r(d) v_n\Big(i - d + y_0, \mathbf{F}(\mathbf{z}), \min\big\{S - i - \sum_{k=0}^{l-1} y_k, q\big\}\Big) \\ &\quad + \big(1 - \mathcal{G}_r^0(i)\big)\, v_n\Big(y_0, \mathbf{F}(\mathbf{z}), \min\big\{S - i - \sum_{k=0}^{l-1} y_k, q\big\}\Big). \end{aligned}$$

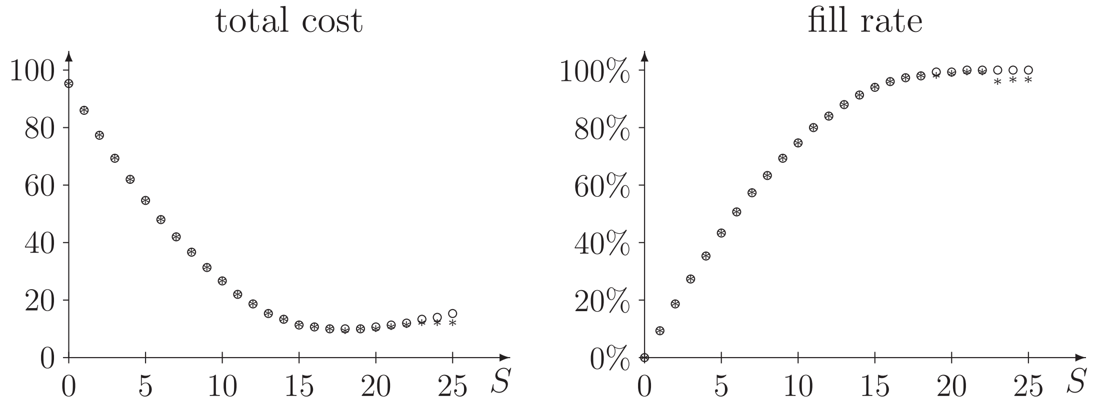


Figure 5.2: The expected total costs and fill rate for different values of S in the PBSP (circle) and RBSP (astrics) where $\lambda = 5$, $\mu = 1$, $R = 1$, $L = 1$, $h = 1$, and $p = 19$.

The one-step transition probabilities from state $(j, \mathbf{z}) \in \mathcal{S}_{s,q}^{t}$ are still specified by Equation (5.2), which provides that Equation (5.7) remains valid as well.

For the RBSP, the values of S and q that minimize the expected total costs in a cost model are denoted by S^* and q^*, respectively. For each base-stock level S, the best value of q is denoted by $q(S)$. This value can be found with a bisection method since all numerical experiments show that the cost function is convex in q for a fixed value of S. Similarly, the value of S^* can be found with a bisection method. Contrary to the PBSP, the fill rate does not have to be an increasing function in S when $q = q(S)$. However, the fill rate is an increasing function in q for a given base-stock level S. Therefore, a bisection method can be used to determine $q(S)$ in a service model, whereas an extensive search procedure is required to find the optimal value of S such that the service level constraint is satisfied.

The PBSP and RBSP are applied to the setting of the previously discussed example for different values of S. In the RBSP, the value of q equals $q(S)$ such that the total costs are minimized for a given base-stock level S. The corresponding average expected total costs and fill rate are shown in Figure 5.2. It illustrates that the cost function is a convex function in base-stock level S. Notice from Figure 5.2 that the total costs in the PBSP decrease with a similar rate as in the RBSP when S increases. However, the costs in the RBSP increase at a lower rate due to the restriction on the maximum order size. Another observation is that the fill rate in the RBSP is not an increasing function in S, since $q(S)$ converges to $E[D_R]$ as S increases. When $q(S+1) < q(S)$ for a base-stock level S, it is possible to order less units of an item when the inventory status is the same at a review even though the base-stock level is larger. This can result in a lower average inventory level and fill rate as illustrated in Figure 5.2 for S equal to 23-25 units.

Optimal values of base-stock level S and maximum order size q are provided in Table 5.3, including the expected total costs and fill rate. A comparison of the order quantities between an optimal policy with the best PBSP and RBSP is illustrated in Table 5.4.

policy	base-stock level	max. order quantity	cost	fill rate
PBSP	18	-	9.77	98.32%
RBSP	18	7	9.66	98.15%

Table 5.3: The best base-stock level and maximum order quantity for the PBSP and RBSP in the example.

i \ y_0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	8	8	8	8	8	8	8	7	6	6	5	4	3	2	1
1	8	8	8	8	8	8	8	7	6	6	5	4	3	2	1
2	8	8	8	8	8	8	7	7	6	5	5	4	3	2	1
3	8	8	8	8	8	8	7	7	6	5	4	3	2	1	
4	8	8	8	8	8	7	7	6	5	5	4	3	2	1	
5	8	8	8	8	7	7	6	6	5	4	3	2	1		
6	8	8	8	7	7	6	6	5	4	3	2	1			
7	8	8	7	7	6	6	5	4	3	2	1				
8	8	7	7	6	6	5	4	3	2	1					
9	7	7	6	6	5	4	3	2	1						
10	7	6	6	5	4	3	2	1							
11	6	6	5	4	3	2	1								
12	6	5	4	3	2	1									
13	5	4	3	2	1										
14	4	3	2	1											
15	3	2	1												
16	2	1													
17	1														

Table 5.4: The order quantities in the PBSP with $S = 18$ coincides with a large part of the optimal policy $a^*(i, y_0)$ (light gray), whereas more quantities are included in the RBSP with $S = 18$ and $q = 7$ (dark gray).

5.3 Approximation model

As discussed in Section 5.1, the numerical computations of the performance measures can be performed by value iteration. This can, however, require quite a computational effort. As shown by Johansen and Thorstenson [138], there are closed-form expressions to compute the value of the performance measures for a lost-sales inventory control system with a PBSP and geometrically distributed demand. Based on these closed-form expressions, we derive a new procedure to approximate the performance measures for inventory systems controlled by a PBSP for any demand distribution and determine near-optimal values of the base-stock level. We extend this approach to the RBSP at the end of this section.

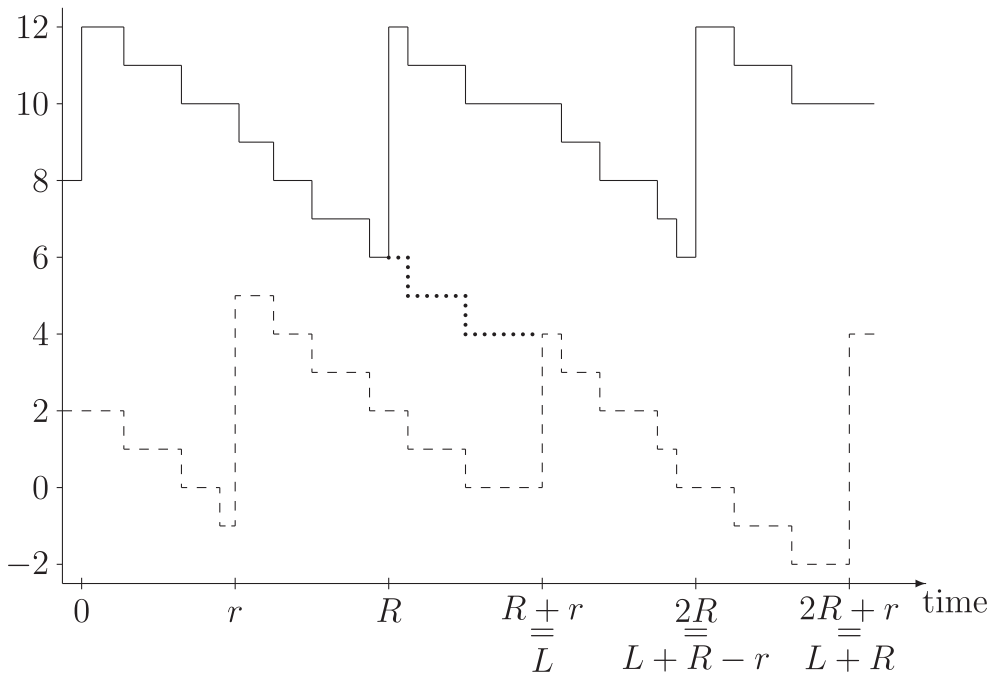
For the PBSP, the inventory position at review time T is equal to base-stock level S. All the orders outstanding at that time arrive before or at time $T+L$. Orders issued after T arrive at time $T+R+L$ or later. Therefore, we consider the time interval $[T+L, T+R+L)$, which is called a shifted review period by Kapalka et al. [154]. In case of backordering, the net inventory at time $\tau \in [T+L, T+R+L)$ equals S minus the demand during the time interval from T to τ. See Figure 5.3 for an illustration of the backorder model where $T=0$. The same setting is used as in the previous example, where $L=1.5R$ and the average demand equals 5 units per unit time. In Figure 5.3, the net inventory level (IL) and inventory position (IP) at time $T=0$ equals 2 and 8 units, respectively. Consequently, there are 6 units on order (O), which were ordered at the previous review and will arrive at time r since the lead time equals $1.5R$. In this example the base-stock level equals 12. Consequently, the inventory position and inventory on order increase both with 4 units at review time $T=0$. The inventory status is also shown below the graphical representation in Figure 5.3. Note that by definition $IP = IL + O$ at all times. This example shows that the on-hand inventory during time period $[T, T+L)$ does not affect the inventory level during $[T+L, T+R+L)$. This is clearly not the case in the lost-sales model as illustrated in Figure 5.4, where the same inventory status and demand occurrences are used as in the backorder system of Figure 5.3.

Because the net inventory cannot be negative in a lost-sales system, we approximate the probability that the on-hand inventory at time $\tau \in [T+L, T+R+L)$ equals i as

$$\tilde{\pi}_\tau(i) = \begin{cases} \tilde{c}_S g_{\tau-T}(S-i), & \text{if } 0 < i \leq S, \\ 1 - \tilde{c}_S \mathcal{G}_{\tau-T}(S), & \text{if } i = 0, \end{cases} \tag{5.14}$$

where the value for factor $\tilde{c}_S$ is derived at the end of this section. Hence, the approximate expected inventory on hand at time $T+L$ is

$$\tilde{IL}_S^+ = \sum_{i=0}^{S} i\tilde{\pi}_{T+L}(i) = \tilde{c}_S E\big[(S-D_L)^+\big] = \tilde{c}_S \mathcal{G}_L^1(S). \tag{5.15}$$



IP	$8 \to 12$	9	$6 \to 12$	10	$6 \to 12$	10
O	$6 \to 10$	$10 \to 4$	$4 \to 10$	$10 \to 6$	$6 \to 12$	$12 \to 6$
IL	2	$-1 \to 5$	2	$0 \to 4$	0	$-2 \to 4$

Figure 5.3: An example on how to determine the on-hand inventory level IL (dashed line) based solely on the inventory position IP (solid line) and the demand during the lead time (dotted line) in a backorder model.

Similarly, the approximation for the expected inventory on hand just before an order delivery at time $T+R+L$ is $\tilde{IL}_S^- = \tilde{c}_S \mathcal{G}^1_{L+R}(S)$. Moreover, the average order size is approximated as $\tilde{IL}_S^+ - \tilde{IL}_S^-$. This number represents the long-run average demand satisfied during a (shifted) review period. As a result, the approximation for the long-run fraction of demand lost is

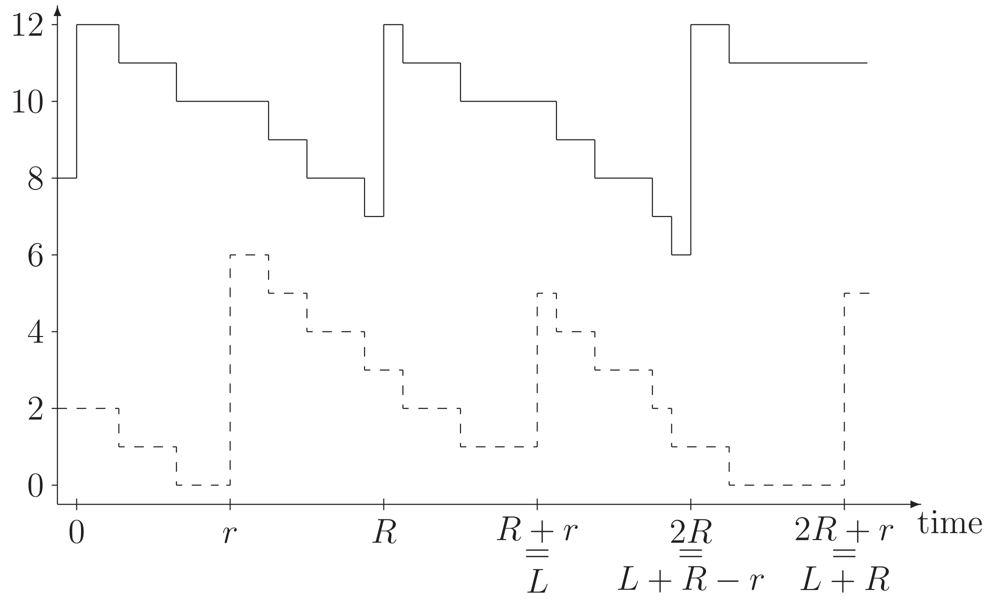
$$\tilde{A}_S = 1 - \tilde{c}_S \frac{\mathcal{G}^1_L(S) - \mathcal{G}^1_{L+R}(S)}{E[D_R]}. \tag{5.16}$$

The approximate average on-hand inventory level is denoted by $\tilde{IL}_S$ and depends on the demand distribution. For a compound Poisson demand process,

$$\tilde{IL}_S = \frac{1}{R} \sum_{i=1}^{S} \tilde{c}_S \big[g_L(S-i) - g_{L+R}(S-i) \big] H(i), \tag{5.17}$$

whereas for a negative binomial demand distribution (or any other demand distribution)

$$\tilde{IL}_S = \frac{\tilde{IL}_S^+ + \tilde{IL}_S^-}{2} = \tilde{c}_S \frac{\mathcal{G}^1_L(S) + \mathcal{G}^1_{L+R}(S)}{2}. \tag{5.18}$$



IP	$8 \to 12$	10	$7 \to 12$	10	$6 \to 12$	11
O	$6 \to 10$	$10 \to 4$	$4 \to 9$	$9 \to 5$	$5 \to 11$	$11 \to 6$
IL	2	$0 \to 6$	3	$1 \to 5$	1	$0 \to 5$

Figure 5.4: The on-hand inventory level IL (dashed line) is not solely based on the inventory position IP (solid line) and the demand in a lost-sales model.

Even though the approximation of Equation (5.18) can be used for any demand distribution, Equation (5.17) is a better approximation in case of a Poisson customer arrival process.

The last aspect of the approximation procedure is to determine $\tilde{c}_S$. We develop two procedures. The first approach is based on the average inventory on order. We apply Little's formula (see Stidham [280]) and equate the approximate average inventory on order with the approximate average amount ordered per unit time multiplied by L. Figure 5.5 illustrates how the inventory on order changes during the shifted review period in the example of Figure 5.4. An order is placed at time $T + L + R - r = T + (l + 1)R$. Consequently, the inventory on order during the interval $[T + (l + 1)R, T + L + R)$ is constant, and it is also constant during the interval $[T + L, T + (l + 1)R)$.

Based on our approximations, the expected inventory on hand at review time $T + L + R - r$ is approximated as $\tilde{c}_S \mathcal{G}^1_{L+R-r}(S)$. Consequently, the average inventory on order during interval $[T + L + R - r, T + L + R)$ is approximated as $S - \tilde{c}_S \mathcal{G}^1_{L+R-r}(S)$. The approximate average inventory on order during $[T + L, T + L + R - r)$ is $\tilde{IL}^+_S - \tilde{IL}^-_S$ units less (i.e., the approximate average order

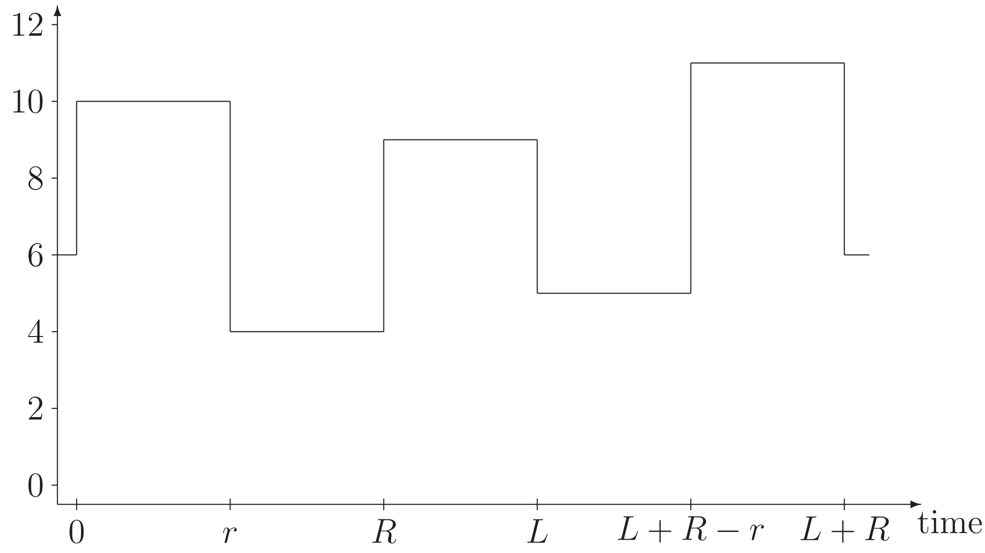


Figure 5.5: The inventory on order O only increases or decreases because of order placements and order deliveries, respectively.

size). Therefore, the average inventory on order is approximated as

$$\begin{aligned}\tilde{O} &= \frac{r}{R}\left[S - \tilde{c}_S\mathcal{G}^1_{L+R-r}(S)\right] + \frac{R-r}{R}\left[S - \tilde{c}_S\mathcal{G}^1_{L+R-r}(S) - \left(\tilde{IL}^+_S - \tilde{IL}^-_S\right)\right] \\ &= S - \tilde{c}_S\left[\mathcal{G}^1_{(l+1)R}(S) + \frac{R-r}{R}\left(\mathcal{G}^1_L(S) - \mathcal{G}^1_{L+R}(S)\right)\right].\end{aligned}$$

Based on Little's formula we set this number equal to $\left(\tilde{IL}^+_S - \tilde{IL}^-_S\right)L/R$, and get

$$\tilde{c}_S = \frac{S}{(l+1)\left(\mathcal{G}^1_L(S) - \mathcal{G}^1_{L+R}(S)\right) + \mathcal{G}^1_{(l+1)R}(S)}. \tag{5.19}$$

A second procedure to set the value of $\tilde{c}_S$ is based on the average order size. As mentioned before, the average order size equals the expected increase of the on-hand inventory level at an order delivery $(\tilde{IL}^+_S - \tilde{IL}^-_S)$. However, the average increase of the inventory position at a review instant should also represent the average order size. This is approximated with the use of Equation (5.14) where $\tau = T + R$. Hence,

$$\tilde{IL}^+_S - \tilde{IL}^-_S = S - \sum_{i=1}^{S} i\tilde{c}_S g_R(S-i) \Leftrightarrow \tilde{c}_S = \frac{S}{\mathcal{G}^1_L(S) + \mathcal{G}^1_R(S) - \mathcal{G}^1_{L+R}(S)}. \tag{5.20}$$

To summarize, the value of base-stock level S which minimizes the approximate expected total costs equal

$$\tilde{S} = \underset{S\geq 0}{\operatorname{argmin}}\left\{h\tilde{IL}_S + pE[D]\tilde{A}_S\right\}, \tag{5.21}$$

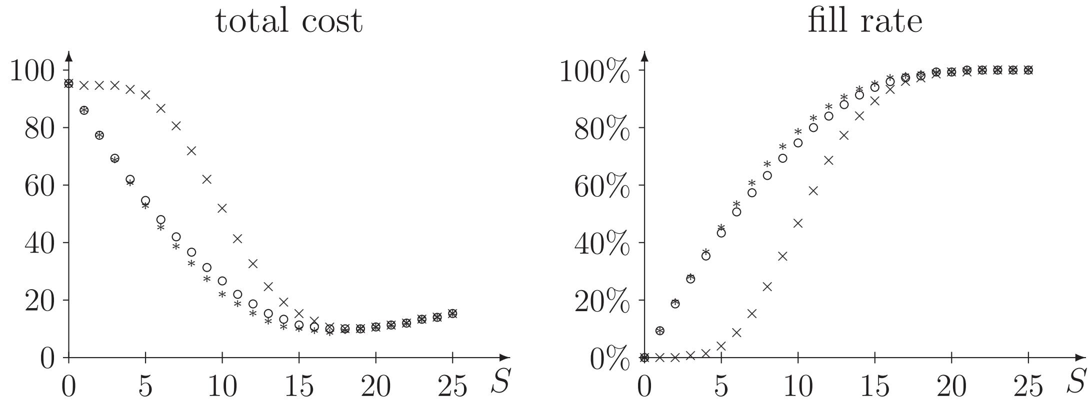


Figure 5.6: The expected total costs and fill rate for different values of base-stock level S in the PBSP when the exact (circle), approximation (astrics) or backorder model (cross) is used where $L = 1.5R$, $\lambda = 5$, $\mu = 1$, $h = 1$, and $p = 19$.

where $\tilde{IL}_S$ and $\tilde{A}_S$ are given by Equation (5.16) to Equation (5.18). The value of S according to Equation (5.21) is denoted by $\tilde{S}_1$ when $\tilde{c}_S$ equals Equation (5.19), and by $\tilde{S}_2$ when $\tilde{c}_S$ equals Equation (5.20). Similarly, for the service model with a minimal service level constraint $\overline{\beta}$

$$\tilde{S} = \min\left\{S \geq 0 \middle| 1 - \tilde{A}_S \geq \overline{\beta}\right\}. \tag{5.22}$$

Notice the resemblance between our approximation procedure for the PBSP in case of lost sales and the performance expressions in case of backordering (Zipkin [327]). When $\tilde{c}_S$ is 1, the expressions derived in this section are exactly those as for the backorder model since no correction has to be made for lost sales. Furthermore, $\lim_{S\to\infty} \tilde{c}_S = 1$ based upon this observation about the backorder model.

To illustrate the performance of our approximation procedure in the analysis of an inventory system with a PBSP, we have computed the exact and approximate expected total costs and fill rate for different base-stock levels. We use the same example as discussed in the previous section. The results are shown in Figure 5.6. Because of very similar results when $\tilde{c}_S$ is based on either of the two approaches in our approximation procedure, we have only included the one where $\tilde{c}_S$ equals Equation (5.19). The results for the backorder model are also included. It appears that the approximation procedure results in an underestimation of the total costs and an overestimation of the fill rate. The opposite is true for the backorder model. More numerical results are discussed in the next section.

In this section, we have discussed a procedure to set the value of base-stock level S based on closed-form expressions that approximate the long-run behavior of a periodic review inventory system with lost sales in case a PBSP is applied. The same base-stock level can be used for the RBSP. However, the value of q should also be specified. As indicated by Figure 5.4, any order placed after time

T is delivered at or after time $T+L+R$. Therefore, the inventory position at time T should cover the demand over $L+R$ time units. Since the inventory position at time T is S, the maximum order size per review period is on average $SR/(L+R)$. Hence, we propose to set the maximum order quantity q equal to $SR/(L+R)$ rounded to the nearest integer (rounded up in case of a tie). The performance of these approximation procedures are analyzed in the next section.

5.4 Numerical results

The goal of this section is twofold. First, we compare the performance of the different replenishment policies for the two demand distributions. Second, we compare different approaches to set suitable values of base-stock level S and maximum order size q.

We illustrate the performance of the (restricted) base-stock policy and optimal policy for a test bed where $R = 1$ and L is ranging from 0.5 to 3.5. Furthermore, the holding cost h equals 1, and the penalty cost p has values of 9, 19, and 39. A compound Poisson process and a negative binomial distribution are used to represent the demand. The parameter values of the demand distribution are set according to Table 5.5.

compound Poisson			
λ	μ	$E[D]$	$Var[D]$
2	1	2	2
5	1	5	5
10	1	10	10
1	2	2	6
2.5	2	5	15
5	2	10	30

negative binomial			
w	u	$E[D]$	$Var[D]$
2	1/2	2	4
2	2/7	5	17.5
10	1/2	10	20

Table 5.5: The parameter values of the demand distributions in our test bed, including the mean ($E[D]$) and variance ($Var[D]$) of the demand per unit time.

Cost model

For the PBSP, we compute the base-stock level $\bar{S}$ that minimizes the expected costs with a bisection method. The value for S that minimizes the approximated costs $\tilde{C}(S)$ where $\tilde{c}_S$ equals Equation (5.19) or Equation (5.20) is denoted by $\tilde{S}_1$ or $\tilde{S}_2$, respectively. The optimal base-stock level in case excess demand is backordered is denoted by S_{BO}. For the RBSP, we compute the base-stock level S^* and maximum order size q^* that minimize the average costs. We have also investigated four other RBSPs with base-stock levels set equal to $\tilde{S}_1$, $\tilde{S}_2$, $\bar{S}$, and S_{BO}. Their maximum order sizes q (denoted $\tilde{q}_1$, $\tilde{q}_2$, $\bar{q}$, and q_{BO}, respectively) are set equal to $SR/(L+R)$ rounded to the nearest integer (rounded up in case of a tie). The intuitive explanation for this choice of q is discussed in Section 5.3. We also report the average costs C^* and fill rate β^* for the optimal replenishment policy, which is computed according to Section 5.2.

Table 5.8 to Table 5.13 provide the results for the pure Poisson, compound Poisson, and negative binomial demand process, respectively. It shows the optimal costs C^*, the parameter values for each of the beforementioned replenishment policies and the corresponding cost increase compared to the costs of an optimal replenishment policy. Table 5.6 gives a summary of these results, where the average is taken over the relative cost increases for each of the policies compared to the optimal policy.

We observe that the base-stock levels found with the approximation procedure where $\tilde{c}_S$ equals Equation (5.19) are smaller than the optimal base-stock levels ($\tilde{S}_1 \leq \bar{S}$). The backorder model, however, finds base-stock levels larger than the optimal values ($\bar{S} \leq S_{\text{BO}}$). This is the case for all test instances, as already observed in the illustration of Figure 5.6 as well. Furthermore, for the approximation procedure where $\tilde{c}_S$ equals Equation (5.20) the following relationship is noticed: $\tilde{S}_1 \leq \tilde{S}_2 \leq S_{\text{BO}}$. The results show that the best RBSP results in policies with almost the same average costs as the optimal policy. We can conclude the same for the RBSP specified by $\bar{S}$ and $\bar{q}$. The results are not decisive whether one of the approximation procedures or the backorder model performs better. In general, the values of $\tilde{S}_1$ perform better than $\tilde{S}_2$. In case of pure Poisson demand, the cost increases in the RBSP with $\tilde{S}_1$ and $\tilde{q}_1$ are irregular compared to the optimal costs, whereas they are on average around 1% for compound Poisson or negative binomial distributed demand.

The results show that the costs deviate relatively more from the optimal costs when the lead time increases. However, this result is less clear when S and q are based on the approximation procedures for the PBSP and RBSP in case of pure Poisson demand. The cost increases in this particular setting vary a lot. For compound Poisson or negative binomial demand, the cost increases can run up to 8% for the RBSP based on the backorder model, while the results for the first approximation procedure are within 2.5% from the optimal solution. Therefore, the base-stock levels found with the approximation model result on average in lower costs compared to the backorder model for the PBSP and RBSP when the demand follows a compound Poisson process or a negative binomial distribution.

We also observe that the cost increases are more significant when the value for q is underestimated compared to the optimal value q^*. Take for instance $L = 1.5$ and $p = 19$ with compound Poisson demand where $\lambda = 5$ and $\mu = 2$. The approximation procedure results in $\tilde{S}_1 = 36$ and $\tilde{q}_1 = 14$, whereas the backorder model results in $S_{\mathrm{BO}} = 40$ and $q_{\mathrm{BO}} = 16$. The best base-stock level and maximum order size are $\bar{S} = S^* = 38$, and $\bar{q} = q^* = 15$, respectively. In the PBSP the base-stock level based on the approximation procedure results in lower costs compared to the best base-stock level in the backorder model. In the RBSP the backorder procedure finds better results than the approximation procedure. This means that the cost function is more sensitive to lower values of q than for higher values.

The results in all tables show that on average the cost increase for using a (restricted) base-stock policy decreases when the penalty cost increases. Furthermore, the cost increase for using $\bar{S}$ and $\bar{q}$ in the RBSP is about the same as the cost increase for using the best values of S and q (compare column 7 and column 8 in Table 5.6). We also observe that the costs for the best RBSP (S^*, q^*) deviate on average 0.5% from the optimal cost, with an observed maximum deviation of 1.35%. The RBSP outperforms the PBSP significantly in all test instances (compare column 3 and column 7 in Table 5.6). Therefore, we conclude that a RBSP performs excellent and should be applied in practice.

		$\bar{S}$	$\tilde{S}_1$	$\tilde{S}_2$	S_{BO}	S^*, q^*	$\bar{S}, \bar{q}$	$\tilde{S}_1, \tilde{q}_1$	$\tilde{S}_2, \tilde{q}_2$	$S_{\mathrm{BO}}, q_{\mathrm{BO}}$
pure	average	1.96%	3.60%	3.52%	3.87%	0.38%	0.43%	2.59%	2.00%	2.15%
Poisson	st.dev.	1.47%	1.93%	2.70%	3.13%	0.31%	0.39%	1.89%	1.94%	2.07%
	max	5.43%	8.07%	10.73%	10.73%	1.35%	1.49%	7.19%	6.59%	7.25%
compound	average	2.06%	2.35%	3.07%	4.12%	0.49%	0.61%	1.05%	1.47%	2.27%
Poisson	st.dev.	1.50%	1.47%	2.45%	3.07%	0.28%	0.41%	0.56%	1.19%	1.80%
	max	5.55%	5.70%	9.48%	13.07%	0.93%	1.72%	2.10%	5.05%	8.06%
negative	average	2.09%	2.44%	3.20%	4.06%	0.50%	0.62%	1.13%	1.52%	2.26%
binomial	st.dev.	1.48%	1.51%	2.64%	3.26%	0.29%	0.43%	0.75%	1.40%	2.09%
	max	5.37%	6.18%	11.18%	13.56%	1.24%	1.79%	2.65%	5.77%	9.51%

Table 5.6: A summary of the results for the cost model: the average, standard deviation and maximum cost increase of each policy compared to the optimal policy.

Service model

The computational effort to solve the service model is much higher compared to the cost model (see Section 5.2.1). Therefore, only a subset of the test bed for the cost model is used as test bed for the service model. We set $h = 1$, $p = 0$, and the minimal fill rate $\overline{\beta}$ equals 85%, 90%, and 95%. The objective is to minimize the average inventory subject to the service level requirement as formulated in Section 5.2. For each problem instance, we compute the optimal and best (restricted) base-stock policy based on the exact lost-sales model, the approximation procedures, and the backorder model. The resulting average on-hand inventory level (IL) and fill rate (β) for each inventory control policy is presented in Table 5.14 to Table 5.17 for Poisson and negative binomial distributed demand. For the optimal policy, we do not report the expected fill rate since the restriction on the fill rate is always satisfied. For the other replenishment policies we report the relative increase of the on-hand inventory level compared to the optimal average inventory level. Based on Figure 5.6 and from the results on the cost model, we already concluded that the approximation procedure where $\tilde{c}_S$ equals Equation (5.19) overestimates the fill rate. Consequently, the resulting base-stock levels found with the approximation procedure are smaller than the optimal base-stock levels ($\tilde{S}_1 \leq \overline{S}$). It is pointless to include the results for the RBSP based on this approximation procedure, since the fill rate is only decreasing when the maximum order size is restricted. For the same reasons we do not present the expected fill rate when the base-stock level or maximum order quantity is based on the backorder model (the service level constraint is always satisfied). A summary is provided in Table 5.7, where the average is taken over the relative increase of the inventory levels for each of the policies compared to the optimal policy. This is only possible for the instances where the service level requirement is met. Therefore, we also include the number of instances in which the requirement is not satisfied. Notice, that there are 27 instances for pure Poisson demand, 9 instances for compound Poisson and 27 instances for negative binomial distributed demand.

Based on these results, we conclude that the average on-hand inventory level in the PBSP is much higher than in the optimal policy ($IL_{\bar{S}} \gg IL^*$). The results for the RBSP are significantly better (compare column 3 and column 7 in Table 5.7). Therefore, we conclude that the RBSP is a good policy for the service model. Finding the optimal values of the base-stock level and maximum order quantity requires an extensive search procedure. It can be reduced when the base-stock level is set to $\bar{S}$ and the maximum order size equals $\bar{S}L/(R+L)$. In that case, the service level restriction is not always satisfied. If it is satisfied, the inventory level increases on average with 5.6% (5.0%) for compound Poisson (negative binomial) distributed demand. When only approximation procedures are required, we recommend to use the PBSP with base-stock level $\tilde{S}_1$. The average cost increase is 4.6% (7.4%) for compound Poisson (negative binomial) distributed demand in case the service level is satisfied. We would not recommend to use the backorder model for the RBSP since the average cost increases are more than 20%, even though the service level constraint is always satisfied.

5.5 Concluding remarks

In this chapter, we considered a lost-sales inventory system with periodic reviews and no fixed order costs. Such inventory systems are commonly seen in a retail environment. A retailer has to decide about the review frequency next to the

		$\bar{S}$	$\tilde{S}_1$	$\tilde{S}_2$	S_{BO}	S^*, q^*	$\bar{S}, \bar{q}$	$\tilde{S}_1, \tilde{q}_1$	$\tilde{S}_2, \tilde{q}_2$	$S_{\text{BO}}, q_{\text{BO}}$
pure	average	13.99%	9.86%	31.03%	41.01%	5.98%	6.53%	8.32%	22.52%	33.09%
Poisson	frequency	-	14	6	-	-	9	21	6	-
compound	average	7.85%	4.58%	18.31%	26.68%	1.79%	2.57%	0.83%	13.69%	19.98%
Poisson	frequency	-	2	1	-	-	4	8	2	-
negative	average	10.25%	7.38%	20.06%	29.94%	3.13%	4.98%	6.38%	13.57%	23.16%
binomial	frequency	-	7	3	-	-	7	17	4	-

Table 5.7: A summary of the results for the service model: the average increase in average inventory level of each policy compared to the optimal policy when the service constraint is met, including the frequency that the service level is not met.

replenishment policy. We developed a general model for inventory systems where lead times and review periods can be of any length. In particular, we studied systems with a cost or service objective, where the total costs consist of the expected holding and penalty costs per unit time. This general structure makes it possible for a retailer to determine the review period length, contrary to the existing literature (see Section 4.7). In practice, pure base-stock policies are often implemented when there is no fixed order cost incurred with an order because of their simplicity. Such replenishment policies are known to be sub-optimal, and can perform bad in case of a low penalty cost for lost sales (Zipkin [328]). Therefore, we proposed a restricted base-stock policy which limits the order size to a maximum number q. From numerical results we concluded that the costs for the best values of S and q deviate on average less than 0.5% from the optimal costs in a cost model, whereas in a service model the average increase of the on-hand inventory level equals 4.2%. Therefore, RBSPs are recommended to be used in practical applications. We also developed an approximation procedure to set near-optimal values of S and q with closed-form expressions for the performance measures such as the total expected costs and fill rate. When there is a cost objective, the results for the approximation procedure show deviations of at most 2.65% from the optimal costs in case demand follows a compound Poisson or negative binomial distribution. Such demand distributions are common to be found in practice.

Base-stock policies are known to perform poorly as replenishment policy for lost-sales inventory systems in which fixed order costs are charged. A reorder level should be included to prevent ordering at each review instant. This is the topic of the next chapter.

λ, μ	L	p	C^*	β^*	$\bar{S}$	$C_{\bar{S}}$	S^*, q^*	$C_{S*,q*}$	$C_{\bar{S},\bar{q}}$
2,1	0.5	9	4.09	93.9%	5	0.77%	5,4	0.19%	0.23%
2,1	0.5	19	4.88	97.7%	6	0.44%	6,4	0.00%	0.00%
2,1	0.5	39	5.62	99.2%	7	0.33%	7,5	0.13%	0.13%
2,1	1.5	9	4.62	91.5%	7	2.01%	7,3	0.13%	0.13%
2,1	1.5	19	5.66	96.1%	8	1.01%	8,3	0.35%	0.35%
2,1	1.5	39	6.58	98.2%	9	0.50%	9,4	0.06%	0.06%
2,1	2.5	9	4.92	90.3%	9	3.61%	9,3	1.35%	1.35%
2,1	2.5	19	6.12	96.1%	11	2.91%	11,3	0.67%	0.67%
2,1	2.5	39	7.20	98.3%	12	1.38%	12,3	0.34%	0.34%
2,1	3.5	9	5.12	89.6%	11	5.22%	11,2	0.69%	0.69%
2,1	3.5	19	6.47	95.2%	13	3.49%	13,3	1.03%	1.03%
2,1	3.5	39	7.73	97.8%	14	1.63%	14,3	0.26%	0.26%
5,1	0.5	9	7.29	96.8%	11	1.18%	11,7	0.15%	0.15%
5,1	0.5	19	8.42	98.5%	12	0.43%	12,8	0.02%	0.02%
5,1	0.5	39	9.43	99.3%	13	0.19%	13,9	0.01%	0.01%
5,1	1.5	9	8.12	95.1%	16	2.42%	16,6	0.15%	0.15%
5,1	1.5	19	9.63	98.1%	18	1.48%	18,7	0.26%	0.26%
5,1	1.5	39	10.99	98.9%	19	0.61%	19,8	0.11%	0.11%
5,1	2.5	9	8.62	94.4%	21	3.92%	21,6	0.59%	0.59%
5,1	2.5	19	10.44	97.6%	23	2.15%	23,6	0.65%	0.65%
5,1	2.5	39	12.05	98.8%	25	1.22%	25,7	0.23%	0.23%
5,1	3.5	9	8.96	93.8%	26	5.43%	27,5	0.89%	1.49%
5,1	3.5	19	11.03	97.2%	28	3.20%	29,6	0.54%	0.83%
5,1	3.5	39	12.88	98.7%	31	2.11%	31,7	0.68%	0.68%
10,1	0.5	9	11.68	97.7%	19	0.91%	19,13	0.16%	0.16%
10,1	0.5	19	13.23	98.8%	21	0.38%	21,15	0.04%	0.07%
10,1	0.5	39	14.57	99.6%	23	0.24%	23,15	0.06%	0.06%
10,1	1.5	9	12.87	96.7%	30	2.31%	30,12	0.28%	0.28%
10,1	1.5	19	14.95	98.7%	32	1.29%	32,13	0.35%	0.35%
10,1	1.5	39	16.76	99.4%	34	0.80%	34,14	0.33%	0.33%
10,1	2.5	9	13.59	96.1%	40	3.73%	40,11	0.37%	0.37%
10,1	2.5	19	16.10	98.3%	43	2.08%	43,12	0.33%	0.33%
10,1	2.5	39	18.30	99.3%	46	1.40%	46,13	0.38%	0.38%
10,1	3.5	9	14.08	95.6%	49	5.02%	50,11	0.83%	1.20%
10,1	3.5	19	16.95	98.0%	53	2.96%	54,11	0.56%	0.84%
10,1	3.5	39	19.50	99.1%	56	1.82%	57,12	0.45%	0.66%

Table 5.8: The results for the exact cost model when the demand follows a pure Poisson process.

λ, μ	L	p	$\tilde{S}_1$	$C_{\tilde{S}_1}$	$\tilde{S}_2$	$C_{\tilde{S}_2}$	S_{BO}	$\text{C}_{S_{\text{BO}}}$	$\text{C}_{\tilde{S}_1,\tilde{q}_1}$	$\text{C}_{\tilde{S}_2,\tilde{q}_2}$	$\text{C}_{S_{\text{BO}},\tilde{q}_{\text{BO}}}$
2,1	0.5	9	5	0.77%	5	0.77%	5	0.77%	0.23%	0.23%	0.23%
2,1	0.5	19	6	0.44%	6	0.44%	6	0.44%	0.00%	0.00%	0.00%
2,1	0.5	39	6	3.09%	6	3.09%	7	0.33%	3.44%	3.44%	0.13%
2,1	1.5	9	7	2.01%	7	2.01%	8	6.09%	0.13%	0.13%	3.16%
2,1	1.5	19	8	1.01%	8	1.01%	9	2.96%	0.35%	0.35%	2.09%
2,1	1.5	39	9	0.50%	9	0.50%	10	2.03%	0.06%	0.06%	1.32%
2,1	2.5	9	9	3.61%	10	6.98%	10	6.98%	1.35%	3.80%	3.80%
2,1	2.5	19	10	3.07%	11	2.91%	11	2.91%	1.62%	0.67%	0.67%
2,1	2.5	39	12	1.38%	12	1.38%	12	1.38%	0.34%	0.34%	0.34%
2,1	3.5	9	11	5.22%	12	8.64%	12	8.64%	0.69%	5.34%	5.34%
2,1	3.5	19	12	4.20%	14	8.04%	14	8.04%	2.41%	4.84%	4.84%
2,1	3.5	39	14	1.63%	15	3.82%	15	3.82%	0.26%	1.58%	1.58%
5,1	0.5	9	10	1.83%	10	1.83%	11	1.18%	1.18%	1.18%	0.15%
5,1	0.5	19	11	3.78%	11	3.78%	12	0.43%	4.73%	4.73%	0.02%
5,1	0.5	39	12	5.05%	12	5.05%	13	0.19%	5.47%	5.47%	0.01%
5,1	1.5	9	15	3.71%	16	2.42%	17	4.72%	2.07%	0.15%	2.67%
5,1	1.5	19	17	2.33%	18	1.48%	18	1.48%	1.47%	0.26%	0.26%
5,1	1.5	39	18	4.26%	19	0.61%	20	1.65%	4.70%	0.11%	0.98%
5,1	2.5	9	20	4.63%	22	5.70%	23	9.74%	1.88%	1.60%	7.25%
5,1	2.5	19	22	3.96%	24	3.27%	25	6.81%	3.18%	1.47%	4.68%
5,1	2.5	39	24	2.88%	26	2.68%	26	2.68%	2.25%	1.27%	1.27%
5,1	3.5	9	25	5.52%	28	10.73%	28	10.73%	2.16%	5.36%	5.36%
5,1	3.5	19	27	5.06%	30	6.10%	30	6.10%	3.24%	3.99%	3.99%
5,1	3.5	39	30	2.19%	32	4.22%	32	4.22%	1.05%	2.49%	2.49%
10,1	0.5	9	18	3.65%	18	3.65%	20	1.20%	3.72%	3.72%	0.24%
10,1	0.5	19	20	2.96%	20	2.96%	22	1.15%	3.49%	3.49%	0.75%
10,1	0.5	39	22	1.24%	22	1.24%	23	0.24%	1.17%	1.17%	0.06%
10,1	1.5	9	28	5.04%	30	2.31%	31	3.47%	4.06%	0.28%	1.07%
10,1	1.5	19	30	7.11%	32	1.29%	33	1.60%	7.19%	0.35%	0.45%
10,1	1.5	39	33	2.98%	34	0.80%	35	0.94%	3.07%	0.33%	0.34%
10,1	2.5	9	37	8.07%	41	4.61%	42	6.56%	5.78%	1.75%	3.38%
10,1	2.5	19	41	4.77%	44	2.84%	45	4.77%	3.60%	1.33%	3.10%
10,1	2.5	39	44	3.13%	46	1.40%	47	2.73%	2.52%	0.38%	1.47%
10,1	3.5	9	47	7.50%	53	10.27%	53	10.27%	5.03%	6.59%	6.59%
10,1	3.5	19	51	5.70%	55	4.48%	56	6.59%	4.64%	1.76%	3.53%
10,1	3.5	39	54	5.25%	58	3.31%	59	5.46%	4.78%	1.90%	3.88%

Table 5.9: The results for the approximate cost model when the demand follows a pure Poisson process.

λ, μ	L	p	C^*	β^*	$\bar{S}$	$C_{\bar{S}}$	S^*, q^*	$C_{S*,q*}$	$C_{\bar{S},\bar{q}}$
1,2	0.5	9	6.95	84.7%	6	0.36%	6,5	0.07%	0.19%
1,2	0.5	19	8.82	92.8%	8	0.21%	8,6	0.03%	0.32%
1,2	0.5	39	10.57	96.7%	10	0.15%	10,8	0.05%	0.08%
1,2	1.5	9	7.57	83.6%	8	1.99%	8,3	0.59%	0.59%
1,2	1.5	19	9.82	92.0%	11	1.43%	11,4	0.49%	0.49%
1,2	1.5	39	11.96	96.0%	13	0.71%	13,6	0.30%	0.35%
1,2	2.5	9	7.93	81.6%	10	3.53%	10,3	0.93%	0.93%
1,2	2.5	19	10.48	91.3%	13	2.23%	13,4	0.75%	0.75%
1,2	2.5	39	12.90	95.7%	16	1.58%	16,4	0.69%	0.70%
1,2	3.5	9	8.16	79.7%	11	4.75%	12,2	0.73%	1.72%
1,2	3.5	19	10.95	90.1%	15	3.15%	15,3	0.90%	0.90%
1,2	3.5	39	13.64	95.3%	18	2.00%	18,4	0.78%	0.78%
2.5,2	0.5	9	11.45	92.2%	13	0.87%	13,9	0.13%	0.13%
2.5,2	0.5	19	13.95	96.7%	16	0.58%	16,11	0.18%	0.18%
2.5,2	0.5	39	16.25	98.3%	18	0.22%	18,13	0.06%	0.10%
2.5,2	1.5	9	12.66	90.6%	18	2.57%	18,7	0.52%	0.52%
2.5,2	1.5	19	15.79	95.6%	22	1.57%	22,8	0.45%	0.46%
2.5,2	1.5	39	18.66	97.9%	25	0.88%	25,10	0.31%	0.31%
2.5,2	2.5	9	13.37	89.2%	23	4.21%	23,6	0.70%	1.26%
2.5,2	2.5	19	16.99	94.9%	27	2.51%	28,7	0.69%	0.83%
2.5,2	2.5	39	20.33	97.7%	31	1.53%	31,9	0.56%	0.56%
2.5,2	3.5	9	13.84	88.1%	27	5.51%	29,5	0.84%	1.37%
2.5,2	3.5	19	17.86	94.4%	32	3.48%	33,7	0.86%	1.07%
2.5,2	3.5	39	21.62	97.4%	37	2.19%	37,8	0.70%	0.70%
5,2	0.5	9	17.25	95.1%	23	1.08%	23,15	0.19%	0.19%
5,2	0.5	19	20.47	97.9%	26	0.47%	26,18	0.10%	0.18%
5,2	0.5	39	23.37	99.0%	29	0.31%	29,21	0.14%	0.22%
5,2	1.5	9	19.11	93.7%	33	2.65%	33,13	0.49%	0.49%
5,2	1.5	19	23.21	97.2%	38	1.50%	38,15	0.37%	0.37%
5,2	1.5	39	26.90	98.7%	42	0.85%	42,17	0.29%	0.29%
5,2	2.5	9	20.21	92.7%	43	4.22%	44,11	0.58%	0.77%
5,2	2.5	19	25.02	96.7%	49	2.51%	49,13	0.58%	0.65%
5,2	2.5	39	29.39	98.5%	54	1.51%	54,15	0.47%	0.47%
5,2	3.5	9	20.94	91.9%	52	5.55%	53,11	0.76%	1.57%
5,2	3.5	19	26.35	96.3%	59	3.40%	60,12	0.68%	0.90%
5,2	3.5	39	31.31	98.3%	65	2.09%	65,14	0.64%	0.64%

Table 5.10: The results for the exact cost model when the demand follows a compound Poisson process.

λ, μ	L	p	$\tilde{S}_1$	$C_{\tilde{S}_1}$	$\tilde{S}_2$	$C_{\tilde{S}_2}$	S_{BO}	$\mathrm{C}_{S_{\mathrm{BO}}}$	$\mathrm{C}_{\tilde{S}_1,\tilde{q}_1}$	$\mathrm{C}_{\tilde{S}_2,\tilde{q}_2}$	$\mathrm{C}_{S_{\mathrm{BO}},\tilde{q}_{\mathrm{BO}}}$
1,2	0.5	9	6	0.36%	6	0.36%	7	1.77%	0.19%	0.19%	1.07%
1,2	0.5	19	8	0.21%	8	0.21%	9	1.24%	0.32%	0.32%	0.85%
1,2	0.5	39	10	0.15%	10	0.15%	10	0.15%	0.08%	0.08%	0.08%
1,2	1.5	9	8	1.99%	8	1.99%	9	3.24%	0.59%	0.59%	1.53%
1,2	1.5	19	11	1.43%	11	1.43%	11	1.43%	0.49%	0.49%	0.49%
1,2	1.5	39	13	0.71%	13	0.71%	14	1.81%	0.35%	0.35%	1.22%
1,2	2.5	9	10	3.53%	11	5.16%	11	5.16%	0.93%	1.65%	1.65%
1,2	2.5	19	13	2.23%	14	3.52%	14	3.52%	0.75%	1.57%	1.57%
1,2	2.5	39	16	1.58%	16	1.58%	17	3.18%	0.70%	0.70%	2.05%
1,2	3.5	9	11	4.75%	13	7.14%	13	7.14%	1.72%	2.89%	2.89%
1,2	3.5	19	15	3.15%	16	4.20%	17	6.69%	0.90%	2.05%	4.08%
1,2	3.5	39	18	2.00%	19	2.81%	19	2.81%	0.78%	1.16%	1.16%
2.5,2	0.5	9	12	1.42%	12	1.42%	14	2.23%	1.21%	1.21%	1.17%
2.5,2	0.5	19	15	0.77%	15	0.77%	16	0.58%	0.64%	0.64%	0.18%
2.5,2	0.5	39	18	0.22%	18	0.22%	19	0.87%	0.10%	0.10%	0.65%
2.5,2	1.5	9	18	2.57%	19	3.27%	20	5.08%	0.52%	1.20%	2.62%
2.5,2	1.5	19	21	1.71%	22	1.57%	23	2.58%	0.94%	0.46%	1.23%
2.5,2	1.5	39	24	1.28%	25	0.88%	26	1.62%	0.88%	0.31%	0.88%
2.5,2	2.5	9	22	4.28%	25	6.57%	26	8.91%	1.34%	2.68%	4.47%
2.5,2	2.5	19	27	2.51%	29	4.07%	30	5.98%	0.83%	1.79%	4.09%
2.5,2	2.5	39	30	1.93%	32	2.03%	33	3.29%	1.15%	0.85%	1.90%
2.5,2	3.5	9	27	5.51%	30	8.37%	31	10.54%	1.37%	4.02%	5.72%
2.5,2	3.5	19	32	3.48%	35	5.94%	36	7.95%	1.07%	3.23%	4.97%
2.5,2	3.5	39	36	2.32%	39	3.92%	40	5.64%	1.10%	2.35%	3.91%
5,2	0.5	9	21	2.22%	21	2.22%	24	1.96%	2.02%	2.02%	0.91%
5,2	0.5	19	25	1.25%	25	1.25%	27	0.72%	1.07%	1.07%	0.26%
5,2	0.5	39	28	1.33%	28	1.33%	30	0.36%	1.32%	1.32%	0.15%
5,2	1.5	9	32	2.91%	34	2.98%	36	5.27%	1.02%	0.82%	2.62%
5,2	1.5	19	36	2.65%	38	1.50%	40	2.82%	2.10%	0.37%	1.53%
5,2	1.5	39	41	1.22%	42	0.85%	44	1.97%	0.88%	0.29%	1.30%
5,2	2.5	9	41	4.76%	46	6.49%	47	8.01%	1.91%	2.70%	3.89%
5,2	2.5	19	47	3.14%	51	3.65%	52	4.82%	1.76%	1.78%	2.79%
5,2	2.5	39	52	2.25%	56	2.62%	57	3.74%	1.54%	1.42%	2.44%
5,2	3.5	9	51	5.70%	57	9.48%	59	13.07%	1.52%	5.05%	8.06%
5,2	3.5	19	57	4.16%	63	6.00%	64	7.39%	2.09%	3.16%	4.37%
5,2	3.5	39	63	2.83%	68	3.72%	69	4.87%	1.77%	1.99%	3.01%

Table 5.11: The results for the approximate cost model when the demand follows a compound Poisson process.

w,u	L	p	C^*	β^*	$\bar{S}$	$C_{\bar{S}}$	S^*,q^*	$C_{S*,q*}$	$C_{\bar{S},\bar{q}}$
2,0.5	0.5	9	5.74	90.2%	6	0.93%	6,4	0.18%	0.18%
2,0.5	0.5	19	7.18	94.5%	7	0.24%	7,5	0.03%	0.03%
2,0.5	0.5	39	8.55	97.9%	9	0.33%	9,6	0.15%	0.15%
2,0.5	1.5	9	6.30	87.0%	8	2.59%	8,3	0.45%	0.45%
2,0.5	1.5	19	8.05	93.9%	10	1.60%	10,4	0.49%	0.49%
2,0.5	1.5	39	9.70	97.3%	11	1.07%	12,5	0.67%	1.09%
2,0.5	2.5	9	6.63	85.3%	9	3.92%	10,3	1.24%	1.79%
2,0.5	2.5	19	8.62	92.9%	12	2.28%	12,3	0.53%	0.35%
2,0.5	2.5	39	10.51	96.5%	14	1.34%	14,4	0.44%	0.44%
2,0.5	3.5	9	6.84	83.7%	11	5.04%	12,2	0.53%	1.31%
2,0.5	3.5	19	9.03	92.3%	14	3.24%	14,3	0.79%	0.79%
2,0.5	3.5	39	11.13	96.4%	16	2.13%	17,3	0.97%	1.05%
2,2/7	0.5	9	12.38	91.1%	13	0.71%	13,9	0.13%	0.13%
2,2/7	0.5	19	15.37	96.0%	16	0.41%	16,12	0.14%	0.18%
2,2/7	0.5	39	18.19	98.0%	19	0.21%	19,14	0.07%	0.11%
2,2/7	1.5	9	13.57	89.3%	18	2.48%	19,7	0.63%	0.63%
2,2/7	1.5	19	17.19	94.9%	22	1.51%	22,9	0.58%	0.58%
2,2/7	1.5	39	20.60	97.6%	26	0.88%	26,11	0.38%	0.42%
2,2/7	2.5	9	14.26	87.9%	23	4.08%	23,6	0.78%	1.24%
2,2/7	2.5	19	18.39	94.2%	28	2.56%	28,8	0.78%	0.78%
2,2/7	2.5	39	22.30	97.3%	32	1.56%	32,9	0.66%	0.66%
2,2/7	3.5	9	14.72	86.7%	27	5.36%	29,5	0.80%	1.38%
2,2/7	3.5	19	19.27	93.6%	33	3.45%	33,7	0.94%	0.94%
2,2/7	3.5	39	23.61	97.0%	38	2.19%	38,8	0.79%	0.79%
10,0.5	0.5	9	14.90	96.1%	21	0.93%	21,15	0.12%	0.15%
10,0.5	0.5	19	17.41	98.2%	24	0.46%	24,17	0.07%	0.09%
10,0.5	0.5	39	19.68	99.3%	27	0.32%	27,18	0.11%	0.11%
10,0.5	1.5	9	16.45	95.1%	32	2.62%	32,12	0.34%	0.50%
10,0.5	1.5	19	19.69	97.8%	35	1.53%	36,14	0.35%	0.53%
10,0.5	1.5	39	22.59	99.0%	39	0.84%	39,16	0.27%	0.27%
10,0.5	2.5	9	17.38	94.2%	41	4.09%	42,11	0.51%	1.07%
10,0.5	2.5	19	21.20	97.4%	46	2.38%	47,13	0.57%	0.62%
10,0.5	2.5	39	24.66	98.8%	50	1.44%	50,14	0.50%	0.50%
10,0.5	3.5	9	18.00	93.5%	51	5.37%	52,11	0.74%	0.91%
10,0.5	3.5	19	22.32	97.0%	57	3.32%	57,12	0.61%	0.99%
10,0.5	3.5	39	26.26	98.6%	61	2.01%	62,13	0.55%	0.73%

Table 5.12: The results for the exact cost model when the demand follows a negative binomial distribution.

w, u	L	p	$\tilde{S}_1$	$C_{\tilde{S}_1}$	$\tilde{S}_2$	$C_{\tilde{S}_2}$	S_{BO}	$\text{C}_{S_{\text{BO}}}$	$\text{C}_{\tilde{S}_1,\tilde{q}_1}$	$\text{C}_{\tilde{S}_2,\tilde{q}_2}$	$\text{C}_{S_{\text{BO}},\tilde{q}_{\text{BO}}}$
2,0.5	0.5	9	6	0.93%	6	0.93%	6	0.93%	0.18%	0.18%	0.18%
2,0.5	0.5	19	7	0.24%	7	0.24%	7	0.24%	0.03%	0.03%	0.03%
2,0.5	0.5	39	9	0.33%	9	0.33%	9	0.33%	0.15%	0.15%	0.15%
2,0.5	1.5	9	8	2.59%	8	2.59%	8	2.59%	0.45%	0.45%	0.45%
2,0.5	1.5	19	10	1.60%	10	1.60%	10	1.60%	0.49%	0.49%	0.49%
2,0.5	1.5	39	11	1.07%	12	1.27%	12	1.27%	1.09%	0.67%	0.67%
2,0.5	2.5	9	9	3.92%	10	4.27%	11	7.70%	1.79%	1.24%	3.65%
2,0.5	2.5	19	12	2.28%	13	4.34%	13	4.34%	0.53%	2.57%	2.57%
2,0.5	2.5	39	14	1.34%	15	2.92%	15	2.92%	0.44%	1.58%	1.58%
2,0.5	3.5	9	11	5.04%	13	9.59%	13	9.59%	1.31%	5.07%	5.07%
2,0.5	3.5	19	14	3.24%	15	4.80%	15	4.80%	0.79%	1.52%	1.52%
2,0.5	3.5	39	16	2.13%	17	2.63%	18	5.29%	1.05%	1.21%	3.52%
2,2/7	0.5	9	13	0.71%	13	0.71%	14	1.45%	0.13%	0.13%	0.58%
2,2/7	0.5	19	16	0.41%	16	0.41%	17	0.73%	0.18%	0.18%	0.35%
2,2/7	0.5	39	19	0.21%	19	0.21%	20	0.51%	0.11%	0.11%	0.33%
2,2/7	1.5	9	18	2.48%	19	2.94%	20	4.34%	0.63%	0.97%	1.99%
2,2/7	1.5	19	22	1.51%	23	1.77%	24	2.91%	0.58%	0.61%	1.74%
2,2/7	1.5	39	26	0.88%	26	0.88%	27	1.34%	0.42%	0.42%	0.72%
2,2/7	2.5	9	22	4.17%	25	6.04%	26	8.01%	1.39%	2.30%	3.75%
2,2/7	2.5	19	27	2.65%	29	3.16%	30	4.41%	1.15%	1.09%	2.53%
2,2/7	2.5	39	32	1.56%	33	1.87%	34	2.80%	0.66%	0.75%	1.68%
2,2/7	3.5	9	27	5.36%	30	7.92%	31	9.82%	1.38%	3.63%	5.07%
2,2/7	3.5	19	32	3.60%	36	6.27%	37	8.17%	1.44%	3.37%	4.98%
2,2/7	3.5	39	37	2.35%	40	3.48%	41	4.83%	1.24%	1.85%	3.02%
10,0.5	0.5	9	20	1.97%	20	1.97%	22	1.35%	1.86%	1.86%	0.40%
10,0.5	0.5	19	23	1.39%	23	1.39%	25	1.03%	1.42%	1.42%	0.57%
10,0.5	0.5	39	25	2.36%	25	2.36%	27	0.32%	2.41%	2.41%	0.11%
10,0.5	1.5	9	30	3.67%	32	2.62%	34	4.95%	2.05%	0.50%	2.85%
10,0.5	1.5	19	34	2.51%	36	1.56%	37	2.45%	1.68%	0.35%	1.27%
10,0.5	1.5	39	37	2.49%	39	0.84%	40	1.49%	2.25%	0.27%	0.82%
10,0.5	2.5	9	40	4.64%	44	6.07%	45	7.84%	1.86%	2.95%	4.46%
10,0.5	2.5	19	45	2.86%	48	3.43%	49	4.84%	1.32%	1.60%	2.83%
10,0.5	2.5	39	49	1.96%	52	2.50%	53	3.91%	1.22%	1.39%	2.67%
10,0.5	3.5	9	49	6.18%	56	11.18%	57	13.56%	2.46%	5.77%	9.51%
10,0.5	3.5	19	55	3.89%	60	6.21%	61	8.02%	1.83%	3.18%	5.69%
10,0.5	3.5	39	59	3.47%	64	4.00%	65	5.55%	2.65%	2.25%	3.64%

Table 5.13: The results for the approximate cost model when the demand follows a negative binomial distribution.

λ, μ	L	β	IL^*	$\bar{S}$	$IL_{\bar{S}}$	$\beta_{\bar{S}}$	S^*, q^*	$IL_{S*,q*}$	β_{S^*,q^*}	$IL_{\bar{S},\bar{q}}$	$\beta_{\bar{S},\bar{q}}$
2,1	0.5	85%	1.99	4	13.97%	87.9%	4,3	9.20%	87.3%	9.20%	87.3%
2,1	0.5	90%	2.49	5	25.39%	94.4%	5,3	18.48%	93.6%	18.48%	93.6%
2,1	0.5	95%	3.26	6	24.42%	97.8%	6,3	16.61%	96.9%	22.55%	97.7%
2,1	1.5	85%	2.28	6	12.17%	86.6%	6,3	7.99%	86.3%	-8.73%	83.0%
2,1	1.5	90%	2.86	7	16.37%	92.3%	7,3	10.69%	91.9%	10.69%	91.9%
2,1	1.5	95%	3.83	8	9.03%	95.9%	8,3	3.21%	95.4%	3.21%	95.4%
2,1	2.5	85%	2.44	8	16.10%	86.5%	8,3	11.17%	86.3%	-5.61%	83.7%
2,1	2.5	90%	3.11	9	13.61%	91.3%	9,3	8.37%	91.0%	8.37%	91.0%
2,1	2.5	95%	4.29	11	20.86%	97.1%	11,3	14.64%	96.7%	14.64%	96.7%
5,1	0.5	85%	3.18	8	16.01%	87.8%	8,5	0.00%	85.0%	0.00%	85.0%
5,1	0.5	90%	3.92	9	13.40%	92.1%	9,6	5.63%	91.2%	5.63%	91.2%
5,1	0.5	95%	5.10	10	3.34%	95.2%	10,8	2.01%	95.2%	0.00%	94.9%
5,1	1.5	85%	3.46	13	23.14%	88.1%	13,5	5.51%	86.2%	5.51%	86.2%
5,1	1.5	90%	4.37	14	12.91%	91.3%	15,5	4.13%	90.6%	5.33%	90.8%
5,1	1.5	95%	5.91	16	9.01%	95.8%	16,6	0.41%	95.1%	0.41%	95.1%
5,1	2.5	85%	3.61	17	16.54%	85.9%	17,6	9.85%	85.6%	0.59%	84.7%
5,1	2.5	90%	4.65	19	15.46%	91.2%	20,5	4.99%	90.9%	-4.01%	89.3%
5,1	2.5	95%	6.35	21	6.60%	95.1%	21,7	3.36%	95.0%	-1.49%	94.6%
10,1	0.5	85%	4.99	14	11.32%	86.5%	14,10	1.19%	85.2%	-6.64%	83.1%
10,1	0.5	90%	6.09	16	13.42%	92.1%	16,11	5.78%	91.3%	5.78%	91.3%
10,1	0.5	95%	7.84	18	7.92%	95.9%	18,12	3.07%	95.5%	3.07%	95.5%
10,1	1.5	85%	5.20	23	15.26%	86.0%	24,9	1.51%	85.3%	-4.06%	83.8%
10,1	1.5	90%	6.52	25	8.72%	90.3%	26,10	2.93%	90.6%	-3.97%	89.1%
10,1	1.5	95%	8.63	28	4.97%	95.0%	28,14	3.72%	95.0%	-3.54%	94.4%
10,1	2.5	85%	5.32	32	19.29%	86.2%	33,9	2.93%	85.6%	-1.89%	84.3%
10,1	2.5	90%	6.76	35	15.64%	91.0%	35,10	1.01%	90.0%	1.01%	90.0%
10,1	2.5	95%	9.16	39	12.77%	95.8%	39,11	3.13%	95.2%	3.13%	95.2%
2.5,2	0.5	85%	5.69	10	2.23%	85.0%	10,10	2.23%	85.0%	-4.01%	83.9%
2.5,2	0.5	90%	7.09	12	5.75%	90.9%	12,8	0.83%	90.1%	0.83%	90.1%
2.5,2	0.5	95%	9.51	15	7.56%	96.0%	15,8	1.59%	95.1%	5.19%	95.7%
2.5,2	1.5	85%	6.46	16	13.95%	86.9%	16,6	1.36%	85.2%	1.36%	85.2%
2.5,2	1.5	90%	8.15	18	9.69%	91.0%	18,7	1.54%	90.1%	1.54%	90.1%
2.5,2	1.5	95%	11.13	21	3.35%	95.2%	21,10	1.33%	95.1%	-1.88%	94.7%
2.5,2	2.5	85%	6.89	21	16.87%	86.6%	21,6	3.92%	85.4%	3.92%	85.4%
2.5,2	2.5	90%	8.90	23	6.88%	90.2%	24,6	1.02%	90.0%	-0.82%	89.5%
2.5,2	2.5	95%	12.22	27	4.36%	95.1%	27,10	2.32%	95.0%	-0.71%	94.7%

Table 5.14: The results for the exact service model when the demand follows a pure or compound Poisson process.

λ, μ	L	β	$\tilde{S}_1$	$IL_{\tilde{S}_1}$	$\beta_{\tilde{S}_1}$	$\tilde{S}_2$	$IL_{\tilde{S}_2}$	$\beta_{\tilde{S}_2}$	S_{BO}	$I_{S_{\text{BO}}}$	$IL_{\tilde{S}_2,\tilde{q}_2}$	$\beta_{\tilde{S}_2,\tilde{q}_2}$	$IL_{S_{\text{BO}},q_{\text{BO}}}$
2,1	0.5	85%	4	14.0%	87.9%	4	14.0%	87.9%	5	56.9%	9.2%	87.3%	48.3%
2,1	0.5	90%	5	25.4%	94.4%	5	25.4%	94.4%	5	25.4%	18.5%	93.6%	18.5%
2,1	0.5	95%	5	-4.0%	94.4%	5	-4.0%	94.4%	6	24.4%	-9.3%	93.6%	22.5%
2,1	1.5	85%	6	12.2%	86.6%	7	45.5%	92.3%	7	45.5%	38.4%	91.9%	38.4%
2,1	1.5	90%	7	16.4%	92.3%	7	16.4%	92.3%	8	46.0%	10.7%	91.9%	38.2%
2,1	1.5	95%	8	9.0%	95.9%	8	9.0%	95.9%	9	32.8%	3.2%	95.4%	31.1%
2,1	2.5	85%	8	16.1%	86.5%	9	44.9%	91.3%	10	77.2%	38.2%	91.0%	68.3%
2,1	2.5	90%	9	13.6%	91.3%	10	38.9%	94.8%	10	38.9%	31.9%	94.4%	31.9%
2,1	2.5	95%	10	0.8%	94.8%	11	20.9%	97.1%	11	20.9%	14.6%	96.7%	14.6%
5,1	0.5	85%	7	-5.5%	82.2%	7	-5.5%	82.2%	9	39.8%	-15.6%	80.4%	30.2%
5,1	0.5	90%	8	-5.9%	87.8%	8	-5.9%	87.8%	10	34.4%	-18.9%	85.0%	30.0%
5,1	0.5	95%	10	3.3%	95.2%	10	3.3%	95.2%	11	20.6%	0.0%	94.9%	16.4%
5,1	1.5	85%	12	5.8%	84.3%	14	42.3%	91.3%	15	63.2%	32.7%	90.8%	51.3%
5,1	1.5	90%	13	-2.3%	88.1%	15	29.6%	93.9%	16	47.6%	20.1%	93.2%	36.0%
5,1	1.5	95%	15	-4.3%	93.9%	16	9.0%	95.8%	17	23.3%	0.4%	95.1%	19.2%
5,1	2.5	85%	16	2.6%	82.6%	20	67.4%	93.4%	20	67.4%	55.3%	92.9%	55.3%
5,1	2.5	90%	18	2.3%	88.7%	21	45.5%	95.1%	21	45.5%	34.4%	94.6%	34.4%
5,1	2.5	95%	21	6.6%	95.1%	23	32.0%	97.6%	23	32.0%	27.8%	97.5%	27.8%
10,1	0.5	85%	12	-12.2%	79.2%	12	-12.2%	79.2%	16	38.4%	-29.0%	74.9%	29.1%
10,1	0.5	90%	14	-8.8%	86.5%	14	-8.8%	86.5%	17	25.8%	-23.5%	83.1%	16.3%
10,1	0.5	95%	16	-11.9%	92.1%	16	-11.9%	92.1%	19	18.8%	-17.9%	91.3%	15.7%
10,1	1.5	85%	21	-3.0%	81.1%	25	36.3%	90.3%	27	60.8%	20.4%	89.1%	48.5%
10,1	1.5	90%	23	-8.1%	86.0%	27	28.2%	93.7%	28	39.0%	18.4%	93.1%	27.7%
10,1	1.5	95%	26	-10.8%	92.1%	29	13.7%	96.2%	30	22.9%	8.4%	96.0%	17.0%
10,1	2.5	85%	29	-3.2%	80.4%	36	57.6%	92.4%	37	69.0%	36.3%	91.2%	55.7%
10,1	2.5	90%	32	-6.1%	86.2%	38	42.6%	94.8%	39	52.8%	30.9%	94.3%	39.7%
10,1	2.5	95%	36	-8.5%	92.4%	41	29.2%	97.3%	42	38.0%	23.0%	97.1%	31.3%
2.5,2	0.5	85%	10	2.2%	85.0%	10	2.2%	85.0%	11	16.8%	-4.0%	83.9%	8.7%
2.5,2	0.5	90%	12	5.7%	90.9%	12	5.7%	90.9%	13	18.2%	0.8%	90.1%	14.6%
2.5,2	0.5	95%	14	-2.2%	94.6%	14	-2.2%	94.6%	15	7.6%	-5.5%	94.2%	5.2%
2.5,2	1.5	85%	15	2.5%	84.3%	17	25.9%	89.1%	18	38.3%	17.0%	88.3%	28.0%
2.5,2	1.5	90%	17	-0.1%	89.1%	19	19.9%	92.6%	20	30.4%	14.2%	92.2%	23.9%
2.5,2	1.5	95%	21	3.3%	95.2%	22	11.4%	96.2%	23	19.7%	7.8%	95.9%	15.8%
2.5,2	2.5	85%	20	7.0%	84.6%	23	38.0%	90.2%	24	49.2%	28.1%	89.5%	38.1%
2.5,2	2.5	90%	23	6.9%	90.2%	25	24.5%	93.0%	26	33.8%	15.0%	92.2%	23.4%
2.5,2	2.5	95%	27	4.4%	95.1%	29	18.7%	96.7%	30	26.2%	12.9%	96.3%	22.3%

Table 5.15: The results for the approximate service model when the demand follows a pure or compound Poisson process.

w,u	L	β	I^*	$\bar{S}$	$IL_{\bar{S}}$	$\beta_{\bar{S}}$	S^*,q^*	$IL_{S*,q*}$	β_{S^*,q^*}	$IL_{\bar{S},\bar{q}}$	$\beta_{\bar{S},\bar{q}}$
2,0.5	0.5	85%	3.16	5	4.68%	85.9%	5,4	2.27%	85.6%	-2.67%	84.3%
2,0.5	0.5	90%	3.94	6	6.57%	91.1%	6,4	3.10%	90.6%	3.10%	90.6%
2,0.5	0.5	95%	5.29	8	14.76%	96.7%	8,4	10.28%	96.1%	12.78%	96.5%
2,0.5	1.5	85%	3.64	8	22.51%	88.9%	8,3	13.25%	87.8%	13.25%	87.8%
2,0.5	1.5	90%	4.58	9	16.17%	92.4%	9,3	6.62%	91.1%	12.37%	92.0%
2,0.5	1.5	95%	6.21	11	15.05%	96.6%	11,3	4.20%	95.3%	10.92%	96.3%
2,0.5	2.5	85%	3.93	10	20.67%	87.9%	11,2	3.65%	85.2%	11.79%	87.1%
2,0.5	2.5	90%	5.03	11	10.20%	91.1%	11,3	1.55%	90.2%	1.55%	90.2%
2,0.5	2.5	95%	6.90	13	5.45%	95.5%	13,4	1.88%	95.2%	1.88%	95.2%
2,2/7	0.5	85%	6.29	11	7.55%	86.3%	11,8	3.07%	85.6%	0.16%	85.0%
2,2/7	0.5	90%	7.92	13	7.26%	91.2%	13,8	1.67%	90.3%	3.68%	90.7%
2,2/7	0.5	95%	10.71	16	5.08%	95.6%	16,9	0.92%	95.1%	3.21%	95.4%
2,2/7	1.5	85%	7.15	16	5.63%	85.2%	16,9	2.86%	85.0%	-5.82%	83.5%
2,2/7	1.5	90%	9.11	19	9.19%	91.1%	20,6	1.67%	90.0%	3.64%	90.5%
2,2/7	1.5	95%	12.51	23	7.56%	95.7%	23,8	1.42%	95.1%	3.52%	95.4%
2,2/7	2.5	85%	7.68	21	8.07%	85.2%	22,6	3.83%	85.6%	-3.84%	83.9%
2,2/7	2.5	90%	9.88	24	6.45%	90.2%	24,10	4.13%	90.1%	-1.80%	89.4%
2,2/7	2.5	95%	13.73	29	7.10%	95.5%	29,8	1.23%	95.1%	1.23%	95.1%
10,0.5	0.5	85%	6.44	15	5.19%	85.1%	16,10	0.84%	85.0%	-7.00%	82.9%
10,0.5	0.5	90%	7.98	18	12.49%	91.9%	18,11	0.85%	90.1%	4.95%	91.0%
10,0.5	0.5	95%	10.57	21	8.60%	96.1%	21,13	2.85%	95.4%	4.83%	95.7%
10,0.5	1.5	85%	6.99	25	14.22%	86.3%	28,9	3.19%	85.5%	-0.42%	84.8%
10,0.5	1.5	90%	8.82	28	12.52%	91.2%	28,11	1.12%	90.1%	1.12%	90.1%
10,0.5	1.5	95%	12.03	32	7.76%	95.6%	32,13	2.28%	95.3%	2.28%	95.3%
10,0.5	2.5	85%	7.25	34	16.72%	85.9%	37,9	3.37%	85.6%	1.26%	84.8%
10,0.5	2.5	90%	9.36	37	8.65%	90.0%	39,10	0.99%	90.2%	-2.21%	89.3%
10,0.5	2.5	95%	12.98	43	10.68%	95.7%	44,11	1.33%	95.1%	2.09%	95.2%

Table 5.16: The results for the exact service model when the demand follows a negative binomial distribution.

w,u	L	β	$\tilde{S}_1$	$IL_{\tilde{S}_1}$	$\beta_{\tilde{S}_1}$	$\tilde{S}_2$	$IL_{\tilde{S}_2}$	$\beta_{\tilde{S}_2}$	S_{BO}	$IL_{S_{\text{BO}}}$	$IL_{\tilde{S}_2,\tilde{q}_2}$	$\beta_{\tilde{S}_2,\tilde{q}_2}$	$IL_{S_{\text{BO}},q_{\text{BO}}}$
2,0.5	0.5	85%	5	4.7%	85.9%	5	4.7%	85.9%	6	32.7%	-2.7%	84.3%	28.4%
2,0.5	0.5	90%	6	6.6%	91.1%	6	6.6%	91.1%	7	30.1%	3.1%	90.6%	28.0%
2,0.5	0.5	95%	8	14.8%	96.7%	8	14.8%	96.7%	8	14.8%	12.8%	96.5%	12.8%
2,0.5	1.5	85%	8	22.5%	88.9%	8	22.5%	88.9%	9	46.0%	13.3%	87.8%	41.2%
2,0.5	1.5	90%	9	16.2%	92.4%	9	16.2%	92.4%	10	35.7%	12.4%	92.0%	31.0%
2,0.5	1.5	95%	10	0.1%	94.8%	11	15.0%	96.6%	11	15.0%	10.9%	96.3%	10.9%
2,0.5	2.5	85%	10	20.7%	87.9%	11	41.1%	91.1%	11	41.1%	30.0%	90.2%	30.0%
2,0.5	2.5	90%	11	10.2%	91.1%	12	27.1%	93.6%	12	27.1%	16.6%	92.6%	16.6%
2,0.5	2.5	95%	13	5.5%	95.5%	14	18.7%	96.9%	14	18.7%	14.7%	96.6%	14.7%
2,2/7	0.5	85%	11	7.6%	86.3%	11	7.6%	86.3%	12	21.1%	0.2%	85.0%	15.4%
2,2/7	0.5	90%	13	7.3%	91.2%	13	7.3%	91.2%	14	18.6%	3.7%	90.7%	14.3%
2,2/7	0.5	95%	16	5.1%	95.6%	16	5.1%	95.6%	17	13.9%	3.2%	95.4%	11.8%
2,2/7	1.5	85%	16	5.6%	85.2%	17	16.4%	87.4%	19	39.0%	8.0%	86.5%	32.0%
2,2/7	1.5	90%	18	0.2%	89.4%	20	18.5%	92.5%	21	28.0%	12.2%	91.9%	21.0%
2,2/7	1.5	95%	22	0.3%	94.8%	23	7.6%	95.7%	24	14.9%	3.5%	95.4%	12.0%
2,2/7	2.5	85%	21	8.1%	85.2%	24	37.0%	90.2%	25	47.4%	26.4%	89.4%	35.6%
2,2/7	2.5	90%	24	6.4%	90.2%	26	22.7%	92.7%	27	31.2%	12.7%	91.9%	24.2%
2,2/7	2.5	95%	28	0.7%	94.7%	30	13.6%	96.2%	31	20.2%	9.5%	96.0%	15.8%
10,0.5	0.5	85%	14	-5.1%	82.2%	14	-5.1%	82.2%	17	27.5%	-20.0%	78.7%	15.5%
10,0.5	0.5	90%	16	-6.4%	87.7%	16	-6.4%	87.7%	19	22.6%	-14.3%	86.4%	16.7%
10,0.5	0.5	95%	19	-7.5%	93.5%	19	-7.5%	93.5%	21	8.6%	-11.9%	93.0%	4.8%
10,0.5	1.5	85%	23	-2.0%	82.3%	27	32.3%	89.7%	29	52.2%	19.6%	88.8%	41.5%
10,0.5	1.5	90%	26	-2.5%	88.1%	29	20.6%	92.5%	31	37.8%	12.2%	92.0%	27.4%
10,0.5	1.5	95%	30	-5.3%	93.7%	32	7.8%	95.6%	34	21.7%	2.3%	95.3%	17.5%
10,0.5	2.5	85%	32	2.9%	82.8%	39	58.1%	92.3%	40	67.5%	41.0%	91.4%	48.8%
10,0.5	2.5	90%	36	2.3%	88.8%	41	37.3%	94.2%	42	45.3%	27.1%	93.7%	34.2%
10,0.5	2.5	95%	41	-1.0%	94.2%	45	23.1%	96.9%	46	29.6%	16.8%	96.7%	22.9%

Table 5.17: The results for the approximate service model when the demand follows a negative binomial distribution.

Chapter 6

Lost sales systems with fixed order costs

In comparison to the previous chapter, inventory systems with fixed order costs are studied in this chapter. Such inventory systems are more general as compared to the models in Chapter 5. The inventory control problems for such systems arise in many practical settings. For instance, when there are costs involved with the actual placement and processing of orders. Each ordered item requires an order pick at a different location in a warehouse, which results in extra costs for each item. Another example where fixed order costs incur are transportation costs. Furthermore, periodic reviews and lost sales are commonly found characteristics in many real-life inventory systems (see also Chapter 1 and Chapter 4). Therefore, our focus is on such systems.

Based on the literature overview in Section 4.3, we concluded that hardly any models have been developed when a fixed cost is incurred for each order in a periodic review inventory system with lost sales. Nahmias [216] develops an exact model to find an optimal replenishment policy. The amount of computational effort to find an optimal policy can be excessive, especially for large lead times (see Section 4.1). Under the assumption that at most one order is outstanding at any time, Hill and Johansen [118] illustrate that an optimal replenishment policy is neither an (R, s, S) nor an (R, s, Q) policy. Moreover, they show that an optimal policy does not have a structure which could be useful to determine optimal order quantities in practical applications. Since (R, s, S) and (R, s, Q) policies are easy to understand and implement they are preferred in practical settings. Such policies are considered by Kapalka et al. [154] and Johansen and Hill [134], respectively. The former authors assume fractional lead times (lead time less than review period length), whereas the latter authors restrict to situations where $s < Q$ (at most one order outstanding at any time). Tijms and Groenevelt [287] study a service model for an (R, s, S) policy. These studies, however, do not compare the results for the optimal control values of such policies to an optimal replenishment policy.

The goal of this chapter is to develop near-optimal replenishment policies that are generally applicable in practice. To achieve this goal, we first extend the existing models to more general lead time and cost circumstances (see Section 4.7 and Chapter 5). In our model we relax the assumption for the lead time to be an integral multiple of the review period length. Furthermore, the computation of the average holding costs is based on the average on-hand inventory level per unit time instead of the inventory level after the demand has occurred in a review period (i.e., at the end of a review period). This is referred to as the *new and more general lead time and cost circumstances* (similar to Chapter 5). Consequently, the length of the lead time and review period can be any number in our model. This is a necessity when the value for either of these two aspects is variable. In our model they are both treated as constant values for which the value is known. However, our model can be embedded in a larger inventory model where optimal values of the lead time or review period length have to be determined. This requires more information about the supply chain, which is out of the scope of this thesis. We focus on replenishment policies for a single-item inventory system at a single echelon level. As mentioned in Section 4.1, an optimal policy is difficult to compute. Furthermore, as we shall illustrate in Section 6.2, the optimal policy is not insightful. Therefore, it is not recommended to be used in practice and near-optimal approximation policies have to be developed that are easy to understand and can be used in real-life inventory systems. Our second contribution is to develop and compare different replenishment policies. Next to the proposed policies from the literature, a new type of replenishment policy is introduced. Besides these exact models, we also derive easy to compute (approximate) expressions to analyze the inventory systems for different replenishment policies such that the inventory control variables can be determined.

As explained, the model in this chapter is closely related to the model developed in Chapter 5. Therefore, a summary of the notation and assumptions is provided in Section 6.1. We also indicate the differences in the models developed in both chapters. The mathematical models for different replenishment policies are formulated in Section 6.2. Besides an optimal replenishment policy, we consider the (R, s, S) policy and (R, s, Q) policy. These policies are referred to as the *pure order-up-to policy* (POUTP) and the *fixed order size policy* (FOSP), respectively. We also introduce a more general type of policies in Section 6.2, in which a maximum order size is imposed in the POUTP. We call this the *restricted order-up-to policy* (ROUTP). As mentioned in Chapter 5, the computation times can increase quite fast when the lead time increases. Therefore, an approximation procedure is derived in Section 6.3 to determine near-optimal values of the inventory control variables in the different replenishment policies. A numerical comparison between the performance of the policies and the approximation procedures is presented in Section 6.4. We are the first authors to perform such a comparison. The conclusions are discussed in Section 6.5.

6.1 Notation and assumptions

The goal of this section is to present the general notation and assumptions, whereas the inventory systems are modeled for the specific replenishment policies in more detail in Section 6.2. Most of the concepts and notation introduced in Section 5.1 can be used to model inventory systems with fixed order costs. The only difference between the two systems is the cost function. Therefore, we only present a summary of the notation and refer to Section 5.1 for more details. All notations to specify the time framework and demand processes are presented in Table 6.1.

R	length of review period
L	length of lead time
r	time between ordering and order delivery within the same review period ($r = L \bmod R$)
l	number of full review periods between ordering and order delivery of the same order ($l = (L-r)/R$ such that $L = lR + r$)
T	time instant at the start of a considered review period (i.e., at a review instant)
t	time instant of a potential order delivery within the same review period as T ($t = T + r$)
D_τ	stochastic variable for the demand during a time period of length τ
$g_\tau(d)$	$P(D_\tau = d)$, $d = 0, 1, \ldots$
$\mathcal{G}^0_\tau(d)$	$P(D_\tau < d) = \sum_{i=0}^{d-1} g_\tau(i)$, $d = 0, 1, \ldots$
$\mathcal{G}^1_\tau(d)$	$E\left[(d - D_\tau)^+\right] = \sum_{i=1}^{d} \mathcal{G}^0_\tau(i)$, $d = 0, 1, \ldots$

Table 6.1: A summary of the notation to specify time and the demand distribution.

	time T	time t
state vector	$(i, \mathbf{y})$	$(j, \mathbf{z})$
inventory level	$i = (j - d_{R-r})^+$	$j = (i - d_r)^+ + y_0$
orders outstanding	$\mathbf{y} = \mathbf{z}$	$\mathbf{z} = (\mathbf{F}(\mathbf{y}), y_l)$
one-step transition probabilities	$P_{(i,\mathbf{y}),(j,\mathbf{F}(\mathbf{y}),y_l)}$	$p_{(j,\mathbf{z}),(i,\mathbf{z})}$
value function	$V_n(i, \mathbf{y})$	$v_n(j, \mathbf{z})$

Table 6.2: Notation to model the inventory system as a Markov chain at a review instant (time T) or order delivery (time t).

Similar to Chapter 5, we consider two types of demand distributions: (compound) Poisson and negative binomial. We assume that demand is independent and identically distributed over time. Furthermore, we denote a set of non-negative integers by $\mathbb{N}_0$, and a set of all integers between m and n by $\mathbb{N}_{m,n} = \{i \in \mathbb{N}_0 \,|\, m \leq i \leq n\} = \{m, m+1, \ldots, n\}$. But also, $\mathbb{N}^{l+1}_{m,n}$ is defined as the $(l+1)$-fold cartesian product of $\mathbb{N}_{m,n}$. This notation is used to model the inventory system as a Markov chain. The state space is an $(l+1)$-dimensional vector space. The first component in the state description prescribes the number of units on hand and the remaining l components specify the order quantities ordered at the previous l review instants. When no order is placed, this component contains the value zero. We also define a function $\mathbf{F}(\mathbf{x})$ which removes the first component of a vector $\mathbf{x}$. An overview of the notation for the Markov chain description is provided in Table 6.2. A distinction is made between the notation at review time T and order delivery time t.

The one-step transition probabilities $P_{(i,\mathbf{y}),(j,\mathbf{F}(\mathbf{y}),y_l)}$ at time T depend on the replenishment policy. The Markov chain can be analyzed with a value iteration algorithm (see Section 5.1). As illustrated in the previous chapter, it is sufficient to formulate a cost function to compute the performance measures of interest, such as the expected total costs, on-hand inventory, or fill rate. Therefore, the value function $V_n(i, \mathbf{y})$ (or $v_n(j, \mathbf{z})$) denotes the total expected costs incurred over a time interval from time T (or t) to time $T + nR$ when the system is in state $(i, \mathbf{y})$ (or $(j, \mathbf{z})$) at time T (or t) and no costs are incurred at and after time $T + nR$. The expected costs over a period of length τ during which no order is delivered or placed is denoted by $c_\tau(i)$ where i represents the amount of on-hand inventory at the beginning of this period. This cost function $c_\tau(i)$ consists of the holding and penalty costs. The average time-weighted on-hand inventory during a period of length τ is denoted by $H_\tau(i)$ for a given initial inventory level of i units, and depends on the demand distribution. For a Poisson customer arrival process with rate λ and a delayed geometric customer demand distribution with mean $\mu = 1/(1-\theta)$,

$$H_\tau(i) = H(i) - \sum_{j=0}^{i-1} g_\tau(j) H(i-j), \tag{6.1}$$

where

$$H(i) = \frac{(i+1)i}{2\lambda} - \theta \frac{(i-1)i}{2\lambda}. \tag{6.2}$$

See for more details Section 5.1 of the previous chapter. Since Equation (6.2) is based on the PASTA property (see Wolff [308]), it is not valid for any other type of demand distribution. For demand distributions with no Poisson customer arrival process, Equation (6.1) is approximated by

$$H_\tau(i) \approx \tau \frac{i + \mathcal{G}_\tau^1(i)}{2}. \tag{6.3}$$

The expected demand exceeding the on-hand inventory of i units during a period of length τ equals

$$E\big[(D_\tau - i)^+\big] = E\left[D_\tau\right] - i + E\big[(i - D_\tau)^+\big] = \tau E[D] - i + \mathcal{G}_\tau^1(i).$$

The unit holding cost per unit time is denoted by h, the unit penalty cost for each lost demand by p, and the fixed order cost by K. Consequently,

$$\begin{aligned} c_\tau(i) &= hH_\tau(i) + pE\big[(D_\tau - i)^+\big], \\ v_n(j, \mathbf{z}) &= c_{R-r}(j) + \sum_{d=0}^{j-1} g_{R-r}(d) V_{n-1}(j-d, \mathbf{z}) \\ &\quad + \left(1 - \mathcal{G}_{R-r}^0(j)\right) V_{n-1}(0, \mathbf{z}) && (6.4) \\ V_n(i, \mathbf{y}) &= c_r(i) + \sum_{j, y_l} P_{(i,\mathbf{y}),(j,\mathbf{F}(\mathbf{y}),y_l)} \Big(K\delta(y_l) + v_n(j, \mathbf{F}(\mathbf{y}), y_l)\Big), && (6.5) \end{aligned}$$

where $\delta(i)$ is zero when $i = 0$ and one otherwise, and $V_0(i, \mathbf{y}) = 0$. Notice that the inclusion of the fixed order cost K in Equation (6.5) is the only difference between the models developed in this chapter and Chapter 5. A value iteration algorithm can be used to compute the expected total costs as discussed in Section 5.1. The value of the order quantity y_l and the transition probabilities $P_{(i,\mathbf{y}),(j,\mathbf{F}(\mathbf{y}),y_l)}$ at a review instant in Equation (6.5) depend on the replenishment policy. Different policies are discussed in the next section.

6.2 Different replenishment policies

The goal of this section is to develop mathematical models for different replenishment policies. We also discuss several properties of the search space to find optimal values of the inventory control variables for the cost model and service model. As we will show, an optimal policy is not easy to derive. Therefore, we also consider other replenishment policies that are easy to implement in practice, such as POUTP, ROUTP, and FOSP. Each of these policies is described in more detail in this section, including the state space and transition probabilities for the Markov decision model as presented in Section 6.1. The main difference between these policies compared to the policies described in Section 5.2 is that a reorder level is included to specify whether or not an order should be placed. This implies that the order size is either zero or restricted to a minimum quantity (i.e., maximum inventory position minus reorder level). Consequently, the state space and transition probabilities of the Markov chain have to be adjusted accordingly.

6.2.1 Optimal policy

To find an optimal replenishment policy and the affiliated expected costs for an inventory system with a cost objective as described in Section 6.1, the state space of the Markov chain is infinite. More specifically, the state space at time T equals $\mathcal{S}^T = \left\{(i, \mathbf{y}) \in \mathbb{N}_0^{1+l}\right\}$ and at time t it is $\mathcal{S}^t = \left\{(j, \mathbf{z}) \in \mathbb{N}_0^{1+l}\right\}$. At time T, the one-step transition probabilities and optimal order quantities depend on the entire state description. Therefore, we use the following value function at time T

$$\begin{aligned} V_n(i, \mathbf{y}) &= c_r(i) + \min_{y_l \geq 0} \Big\{ K\delta(y_l) + \sum_{d=0}^{i-1} g_r(d) v_n(i - d + y_0, \mathbf{F}(\mathbf{y}), y_l) \\ &\qquad + \left(1 - \mathcal{G}_r^0(i)\right) v_n(y_0, \mathbf{F}(\mathbf{y}), y_l) \Big\}. \end{aligned} \tag{6.6}$$

The value function at time t is specified by Equation (6.4). A value iteration algorithm can be used to solve this Markov decision problem. The same solution procedure as in Section 5.2 can be used to solve the service model. An upper bound on the state space would improve the efficiency of such solution approaches. To derive such a bound, we relate the backorder system to a lost-sales system. Since lost sales do not decrease the inventory position, the optimal order quantities in lost-sales systems are never higher compared to backorder systems. The optimal replenishment policy for a backorder system is a pure order-up-to policy. Therefore, the inventory position in a lost-sales system would be less than or equal to the optimal order-up-to level for a similar inventory system with backorders. There is however no proof for this intuitive upper bound on the state space. The computation of this upper bound is discussed in Section 6.3 on backorder models.

To illustrate the structure of an optimal policy for a cost model, we consider the same example as in Section 5.2 with the extension that fixed order costs are charged. In this example, $R = 1$, $L = 1.5$, $h = 1$, $p = 19$ and $K = 50$. The customer's demand is a pure Poisson process with $\lambda = 5$ and $\mu = 1$. Table 6.3

presents the optimal order quantities $a^*(i, y_0)$ when the on-hand inventory level equals i units at a review instant (first column) and the inventory on order is y_0 (first row). The corresponding expected total costs are 27.34 and the fill rate equals 97.39%. When we compare this optimal policy to the optimal policy when $K = 0$ (Table 5.2), we see that the order quantities are larger. Furthermore, the order quantities are either zero, or larger than or equal to 22 units. This indicates that reorder levels should be introduced. We also notice that the optimal order quantities are decreasing in the inventory position and the rate of decrease is less than one. This has been observed before for periodic review systems (Section 5.2) and continuous review systems (Section 4.2).

i \ y_0	0	1	2	3	4	5	6	7	8	9
0	27	27	27	27	26	26	25	25	24	23
1	27	27	27	27	26	26	25	25	24	23
2	27	27	27	27	26	26	25	24	23	22
3	27	27	27	26	26	25	25	24	23	22
4	27	27	26	26	25	25	24	23	22	
5	27	26	26	25	25	24	23	22		
6	26	26	25	25	24	23	22			
7	26	25	25	24	23	22				
8	25	25	24	23	22					
9	25	24	23	22						
10	24	23	22							
11	23	22								
12	22									

Table 6.3: The optimal order quantities $a^*(i, y_0)$ when the on-hand inventory level equals i and the outstanding order has size y_0 at a review instant.

As mentioned before, Hill and Johansen [118] have developed a policy iteration algorithm to find optimal order quantities when it is assumed that at most one order is outstanding at any time. When we apply their solution procedure to our example, Table 6.4 provides the order quantities $a(i, y_0)$ when $y_0 = 0$, whereas $a(i, y_0) = 0$ when $y_0 > 0$. The expected total costs of this policy are 27.34 and the fill rate equals 97.39%. Notice that the order quantities are the same as the optimal order quantities when $y_0 = 0$ (column 2 in Table 6.4). Surprisingly, the performance of this policy is the same as for the optimal policy. Hence, the assumption of at most one order outstanding at any time seems justified in this example. As we shall see in Section 6.4, this is not always the case.

i	0	1	2	3	4	5	6	7	8	9	10	11	12
$a(i,0)$	27	27	27	27	27	27	26	26	25	25	24	23	22

Table 6.4: The optimal order quantities $a(i,0)$ when at most one order may be outstanding at any time. Hence, $a(i,y_0)=0$ when $y_0>0$.

6.2.2 Pure order-up-to policy (POUTP)

The POUTP with reorder level s and order-up-to level S prescribes to issue a replenishment order at a review time T when the inventory position is at or below reorder level s. The order size is such that the inventory position is raised to level S. This corresponds to (R,s,S) policies. Hence,

$$y_l = \begin{cases} 0, & \text{if } i + \sum_{k=0}^{l-1} y_k > s, \\ S - i - \sum_{k=0}^{l-1} y_k, & \text{otherwise.} \end{cases}$$

The state space of the Markov chain at time T is $\mathcal{S}^T_{s,S} = \bigcup_{i=0}^{S} \big(\{i\} \times \mathbb{Y}_{s,S}(i)\big)$, where

$$\mathbb{Y}_{s,S}(i) = \left\{ \mathbf{y} \in \mathbb{M}^l_{s,S} \,\middle|\, \sum_{k=0}^{l-1} y_k \leq S - i \right\},$$

and $\mathbb{M}_{s,S} = \{0\} \bigcup \mathbb{N}_{S-s,S}$. The state space at time t equals $\mathcal{S}^t_{s,S} = \bigcup_{j=0}^{S} \big(\{j\} \times \mathbb{Z}_{s,S}(j)\big)$, where

$$\mathbb{Z}_{s,S}(j) = \left\{ \mathbf{z} \in \mathbb{M}^l_{s,S} \,\middle|\, \sum_{k=1}^{l} z_k \leq S - j \right\}.$$

The one-step transition probabilities from state $(i,\mathbf{y}) \in \mathcal{S}^T_{s,S}$ with $i + \sum_{k=0}^{l-1} y_k \leq s$ at time T are

$$P_{(i,\mathbf{y}),(i-d+y_0,\mathbf{F}(\mathbf{y}),S-i-\sum_{k=0}^{l-1} y_k)} = \begin{cases} g_r(d), & \text{if } 0 \leq d < i, \\ 1 - \mathcal{G}^0_r(i), & \text{if } d = i, \\ 0, & \text{otherwise,} \end{cases} \tag{6.7}$$

whereas for state $(i,\mathbf{y}) \in \mathcal{S}^T_{s,S}$ with $i + \sum_{k=0}^{l-1} y_k > s$

$$P_{(i,\mathbf{y}),(i-d+y_0,\mathbf{F}(\mathbf{y}),0)} = \begin{cases} g_r(d), & \text{if } 0 \leq d < i, \\ 1 - \mathcal{G}^0_r(i), & \text{if } d = i, \\ 0, & \text{otherwise.} \end{cases} \tag{6.8}$$

According to Equation (6.5) we have for all $(i,\mathbf{y}) \in \mathcal{S}_{s,S}^T$ with $i + \sum_{k=0}^{l-1} y_k \leq s$

$$\begin{aligned} V_n(i,\mathbf{y}) &= K + c_r(i) + \sum_{d=0}^{i-1} g_r(d) v_n\Big(i - d + y_0, \mathbf{F}(\mathbf{y}), S - i - \sum_{k=0}^{l-1} y_k\Big) \\ &\quad + \big(1 - \mathcal{G}_r^0(i)\big)\, v_n\Big(y_0, \mathbf{F}(\mathbf{z}), S - i - \sum_{k=0}^{l-1} y_k\Big), \end{aligned} \tag{6.9}$$

and for all $(i,\mathbf{y}) \in \mathcal{S}_{s,S}^T$ with $i + \sum_{k=0}^{l-1} y_k > s$

$$\begin{aligned} V_n(i,\mathbf{y}) &= c_r(i) + \sum_{d=0}^{i-1} g_r(d) v_n\Big(i - d + y_0, \mathbf{F}(\mathbf{y}), 0\Big) \\ &\quad + \big(1 - \mathcal{G}_r^0(i)\big)\, v_n\Big(y_0, \mathbf{F}(\mathbf{z}), 0\Big), \end{aligned} \tag{6.10}$$

whereas Equation (6.4) specifies $v_n(j,\mathbf{z})$ for all $(j,\mathbf{z}) \in \mathcal{S}_{s,S}^t$.

To solve the cost model when a POUTP is applied, the values of reorder level s and order-up-to level S have to be found such that the expected total costs are minimized. Let us denote these values by $\overline{s}$ and $\overline{S}$, respectively. Numerical results indicate that the cost function is convex in s for a given order-up-to level S. If this would be a property of the cost function, then the optimal value of reorder level s can be determined with a bisection method for each order-up-to level S. Let us denote this value by $s(S)$. Moreover, the cost function seems to be convex in S when the reorder level equals $s(S)$. Even tough we are not able to prove these statements, they are satisfied in all numerical results (see Section 6.4). Therefore, a bisection method is proposed to determine the best order-up-to level $\overline{S}$. For a service model, the value of $s(S)$ can be found by the smallest reorder level such that the service level restriction $\overline{\beta}$ is satisfied, which can be found with a bisection method based on similar reasons. Because of the restriction on the service level, convexity of the cost objective in S cannot be assumed when the reorder level equals $s(S)$. Therefore, the best order-up-to level is determined with an extensive search procedure in the service model.

6.2.3 Restricted order-up-to policy (ROUTP)

The ROUTP with reorder level s, order-up-to level S and upper limit q on the order quantity prescribes to issue a replenishment order at a review time T where the size of the order equals

$$y_l = \begin{cases} 0, & \text{if } i + \sum_{k=0}^{l-1} y_k > s, \\ S - i - \sum_{k=0}^{l-1} y_k, & \text{if } S - q < i + \sum_{k=0}^{l-1} y_k \leq s, \\ q, & \text{otherwise.} \end{cases}$$

The state space of the Markov chain at time T is $\mathcal{S}^T_{s,S,q} = \bigcup_{i=0}^{S} \left(\{i\} \times \mathbb{Y}_{s,S,q}(i)\right)$, where

$$\mathbb{Y}_{s,S,q}(i) = \left\{ \mathbf{y} \in \mathbb{M}^l_{s,S,q} \middle| \sum_{k=0}^{l-1} y_k \leq S - i \right\},$$

and $\mathbb{M}_{s,S,q} = \{0\} \bigcup \{\mathbb{N}_{0,q} \bigcap \mathbb{M}_{s,S}\}$. The state space at time t equals $\mathcal{S}^t_{s,S,q} = \bigcup_{j=0}^{S} \left(\{j\} \times \mathbb{Z}_{s,S,q}(j)\right)$, where

$$\mathbb{Z}_{s,S,q}(j) = \left\{ \mathbf{z} \in \mathbb{M}^l_{s,S,q} \middle| \sum_{k=1}^{l} z_k \leq S - j \right\}.$$

Due to the restriction on the order size y_l in the RBSP, the value function for all $(i, \mathbf{y}) \in \mathcal{S}^T_{s,S,q}$ with $i + \sum_{k=0}^{l-1} y_k \leq s$ equals

$$\begin{aligned} V_n(i, \mathbf{y}) &= K + c_r(i) + \sum_{d=0}^{i-1} g_r(d) v_n\Big(i - d + y_0, \mathbf{F}(\mathbf{y}), \min\Big\{S - i - \sum_{k=0}^{l-1} y_k, q\Big\}\Big) \\ &\quad + \left(1 - \mathcal{G}^0_r(i)\right) v_n\Big(y_0, \mathbf{F}(\mathbf{z}), \min\Big\{S - i - \sum_{k=0}^{l-1} y_k, q\Big\}\Big), \end{aligned}$$

whereas Equation (6.10) and Equation (6.4) specify the value function for all $(i, \mathbf{y}) \in \mathcal{S}^T_{s,S,q}$ with $i + \sum y_k > s$ and for all $(j, \mathbf{z}) \in \mathcal{S}^t_{s,S,q}$, respectively.

For the ROUTP, the values of s, S, and q which minimize the expected average total costs are denoted by s^*, S^* and q^*, respectively. Because of the restriction on the maximum order size, the cost function does not have any convexity properties. Consequently, extensive numerical search procedures are required to determine the optimal values of s^*, S^* and q^*.

6.2.4 Fixed order size policy (FOSP)

In the FOSP, orders of a fixed size Q are placed when the inventory position is at or below reorder level s. This corresponds to (R, s, Q) policies. Hence,

$$y_l = \begin{cases} 0, & \text{if } i + \sum_{k=0}^{l-1} y_k > s, \\ Q, & \text{otherwise.} \end{cases}$$

The state space for the Markov chain of the FOSP is much more compact, since the order quantities are either 0 or Q. The state space at time T equals $\mathcal{S}^T_{s,Q} = \bigcup_{i=0}^{s+Q} \big(\{i\} \times \mathbb{Y}_{s,Q}(i)\big)$, where

$$\mathbb{Y}_{s,Q}(i) = \left\{ \mathbf{y} \in \mathbb{M}^l_Q \,\middle|\, i + \sum_{k=0}^{l-1} y_k \leq s + Q \right\},$$

and $\mathbb{M}_Q = \{0\} \bigcup \{Q\}$. Moreover, at time t the state space equals $\mathcal{S}^t_{s,Q} = \bigcup_{j=0}^{s+Q} \big(\{j\} \times \mathbb{Y}_{s,Q}(i)\big)$. The same structure as in the POUTP and ROUTP is used to specify the value function. For all $(i, \mathbf{y}) \in \mathcal{S}^T_{s,Q}$ with $i + \sum_{k=0}^{l-1} y_k \leq s$

$$\begin{aligned} V_n(i, \mathbf{y}) \;&=\; K + c_r(i) + \sum_{d=0}^{i-1} g_r(d) v_n\Big(i - d + y_0, \mathbf{F}(\mathbf{y}), Q\Big) \\ &\quad + \big(1 - \mathcal{G}^0_r(i)\big)\, v_n\Big(y_0, \mathbf{F}(\mathbf{z}), Q\Big), \end{aligned}$$

and for all $(i, \mathbf{y}) \in \mathcal{S}^T_{s,Q}$ with $i + \sum_{k=0}^{l-1} y_k > s$ the value function $V_n(i, \mathbf{y})$ is specified by Equation (6.10).

For the FOSP, the values of s and Q that minimize the expected average total costs are denoted $\hat{s}$ and $\hat{Q}$, respectively. For each value of fixed order size Q, the best value of reorder level s is denoted $s(Q)$. The value of $s(Q)$ can be found based on a bisection method for each value of Q based on similar reasons as for the POUTP (see Section 6.2.2). Moreover, we assume that the cost function is convex in Q when the reorder level equals $s(Q)$. Consequently, a bisection approach can be used to find the optimal value of Q. Notice that this solution method can result in a local optimum, since we are not able to prove this convexity property of the cost function. However, in all numerical results the global optimum is found with such a solution approach (see Section 6.4). In a service model, the value of $s(Q)$ is the smallest value of reorder level s such that the service level constraint is satisfied for a given value of Q. Similar to the POUTP, a bisection method can be used to find $s(Q)$, whereas an extensive search procedure is required to determine an optimal value of Q.

We compare the three replenishment policies in the remainder of this section based on the same example as mentioned before. Notice that when $q \geq S$ in the ROUTP, the policy corresponds to the POUTP and when $S \geq s + Q$ it represents the FOSP with $Q = q$. Therefore, the ROUTP is a more general class of replenishment policies, and will always outperform the other two policies. For each policy the inventory system is analyzed for different values of S or Q depending on the replenishment policy. The values of the reorder levels and maximum order sizes are set such that they minimize the expected total costs for a given value of S or Q. The corresponding expected total costs and fill rate are shown in Figure 6.1. This example illustrates that the performance for the POUTP and ROUTP is rather similar. Furthermore, the optimal costs for the FOSP are almost the same as for the POUTP and ROUTP. Therefore, all three replenishment policies result in similar optimal costs in this example. Figure 6.1 also illustrates the convexity property of the cost function. An overview on the optimal values of the inventory control variables in a cost model is provided in Table 6.5. The relative cost increase compared to an optimal replenishment policy is also included. It illustrates that the best ROUTP performs the same as an optimal replenishment policy. Therefore, we also include a comparison between the order quantities for an optimal replenishment policy with the best POUTP and ROUTP in Table 6.6. Furthermore, it is interesting to see that the optimal values of the reorder level and order-up-to level are equal for the POUTP and ROUTP. This is however not generally true for any inventory system. More numerical results are discussed in Section 6.4.

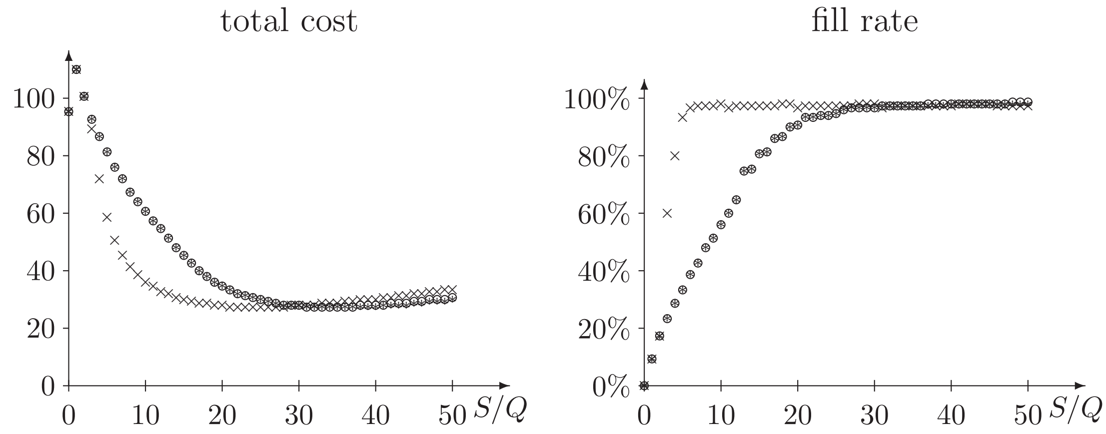


Figure 6.1: The expected total costs and fill rate for different values of S in the POUTP (circle), ROUTP (astrics), and of Q in the FOSP (cross) where $\lambda = 5$, $\mu = 1$, $R = 1$, $L = 1$, $h = 1$, $p = 19$, and $K = 50$.

policy	s	S	q	cost	fill rate	cost increase
optimal	-	-	-	27.34	97.39%	-
POUTP	12	34	(34)	27.36	97.43%	0.05%
ROUTP	12	34	26	27.34	97.39%	0.00%
FOSP	12	(36)	24	27.38	97.38%	0.15%

Table 6.5: The optimal values of the inventory control variables for the different replenishment policies and the corresponding expected total costs and fill rate.

i \ y_0	0	1	2	3	4	5	6	7	8	9
0	27	27	27	27	26	26	25	25	24	23
1	27	27	27	27	26	26	25	25	24	23
2	27	27	27	27	26	26	25	24	23	22
3	27	27	27	26	26	25	25	24	23	22
4	27	27	26	26	25	25	24	23	22	
5	27	26	26	25	25	24	23	22		
6	26	26	25	25	24	23	22			
7	26	25	25	24	23	22				
8	25	25	24	23	22					
9	25	24	23	22						
10	24	23	22							
11	23	22								
12	22									

Table 6.6: The order quantities in the POUTP with $s = 12$ and $S = 34$ coincide with a large part of the optimal policy $a^*(i, y_0)$ (light gray), whereas more quantities are included in the ROUTP with $s = 12$, $S = 34$, and $q = 26$ (dark gray).

6.3 Approximation model

As described in the previous section, it requires multi-dimensional bisection approaches and extensive numerical search procedures to find optimal values of the inventory control variables. In each step of such procedures the corresponding $(l + 1)$-dimensional Markov decision problem has to be solved (see Section 6.2). This can however require a lot of computational effort, especially for large lead times (see Section 4.1). The goal of this section is to find near-optimal values of the control variables based on approximate expressions for the performance measures of interest, such as the expected average total costs and fill rate. Johansen and Hill [134] already have proposed an approximation procedure to determine near-optimal values of s and Q for the FOSP. The authors assume that at most one order is outstanding at any time. The performance of this procedure is tested in Section 6.4. In this section, we derive an approximation procedure for a lost-sales inventory system with a POUTP. At the end of this section, we also discuss how this approximation procedure can be used for ROUTPs.

The basic thought behind the approximation procedure is to adjust a back-order model with a correction factor for lost demands similar to Section 5.3. It is common in lost-sales inventory literature to use the backorder system as approximation for the lost-sales system (see Chapter 4). The approximation procedure consists of several steps. First, we derive approximate expressions for the equilibrium distribution of the inventory position and on-hand inventory level. Next, we use these distributions to develop expressions to analyze the performance of the inventory system. In the final step the value of the correction factor is derived. This approach corresponds to the methodology discussed in Section 4.1.

As in the approximation procedure for a pure base-stock policy (see Section 5.3), we relate the on-hand inventory level at an order delivery to the inventory position at a review instant. However, in a base-stock policy there is no reorder level included. Consequently, the inventory position always equals the order-up-to level S at a review instant. This is not the case in order-up-to policies since no order is placed when the inventory position is larger than the reorder level at a review instant. Therefore, let $IP(T)$ and $IL(t)$ be the inventory position at review time T and the net inventory level at order delivery time t, respectively. The superscript '+' indicates that an order has already been placed or delivered at time T or t, respectively. When this is not the case, it is denoted by superscript '-'. Furthermore, $D(T, T+L)$ denotes the demand during the time period $(T, T+L]$. Notice that all the orders outstanding at time T are delivered before or at time $T+L$. Thus all the units in the inventory position $IP^+(T)$ contribute to the inventory level $IL^+(T+L)$. The demand in period $(T, T+L]$, however, depletes this inventory level. In the backorder model, we have the following relationship (see also Figure 5.3),

$$IL^+(T+L) = IP^+(T) - D(T, T+L),$$

whereas in the lost-sales model this is not true (see Figure 5.4 as counterexample). Next, we derive approximate expressions to describe the steady-state behavior of $IP(T)$ and $IL(t)$, denoted by IP and IL, respectively.

In the POUTP, when the inventory position reaches or drops below reorder level s, an order is placed at the next review instant such that the inventory position is raised to S. Hence, the inventory position is always between $s+1$ and S at a review instant, i.e., $IP^+(T) \in \mathbb{N}_{s+1,S}$. To find the steady-state distribution of the inventory position after ordering at time T, we first consider the backorder model. When excess demand is backordered,

$$IP^-(T+R) = IP^+(T) - D(T, T+R),$$

and

$$IP^+(T+R) = \begin{cases} S, & \text{if } IP^-(T+R) \leq s, \\ IP^-(T+R), & \text{otherwise.} \end{cases}$$

Consequently, the inventory position after ordering can be modeled as a one-dimensional Markov chain with state space $\mathbb{N}_{s+1,S}$ in the backorder model. The one-step transitions between the states in this Markov chain are graphically represented in Figure 6.2. The stationary probabilities for $IP^+(T)$ are denoted by π^+_{IP} and given by the solution of

$$\begin{aligned} \pi^+_{IP}(S) &= g_R(0)\pi^+_{IP}(S) + \sum_{i=s+1}^{S} \left(1 - \mathcal{G}^0_R(i-s)\right) \pi^+_{IP}(i), \\ \pi^+_{IP}(j) &= \sum_{i=j}^{S} g_R(i-j)\pi^+_{IP}(i), \qquad j = s+1, \ldots, S-1, \\ \sum_{j=s+1}^{S} \pi^+_{IP}(j) &= 1. \end{aligned} \tag{6.11}$$

These expressions for the backorder model are used to derive approximations for the steady-state distribution of the inventory position IP^+ in the lost-sales model. We adjust the distribution for the demand during the review period in Equation (6.11) with a correction factor $\tilde{c}_{s,S}$ such that the inventory position only

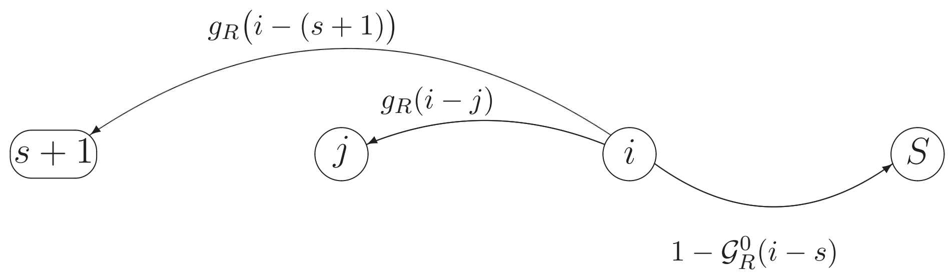


Figure 6.2: The one-step transitions including their probabilities of the inventory position after ordering at a review in the backorder model.

decreases when demand is satisfied. Hence, for the lost-sales model

$$\tilde{\pi}^+_{IP}(S) = \tilde{c}_{s,S} g_R(0) \tilde{\pi}^+_{IP}(S) + \sum_{i=s+1}^{S} \left(1 - \tilde{c}_{s,S} \mathcal{G}^0_R(i-s)\right) \tilde{\pi}^+_{IP}(i), \quad (6.12)$$

$$\tilde{\pi}^+_{IP}(j) = \sum_{i=j}^{S} \tilde{c}_{s,S} g_R(i-j) \tilde{\pi}^+_{IP}(i), \qquad j = s+1, \dots, S-1, \quad (6.13)$$

$$\sum_{j=s+1}^{S} \tilde{\pi}^+_{IP}(j) = 1, \quad (6.14)$$

where $\tilde{\pi}^+_{IP}$ is the approximate steady-state behavior of IP^+. It can easily be verified that the solution to this set of equations is

$$\tilde{\pi}^+_{IP}(j) = \frac{f(S-j)}{\sum\limits_{i=s+1}^{S} f(S-i)}, \qquad j = s+1, \dots, S,$$

where

$$f(j) = \begin{cases} 1, & \text{if } j = 0, \\ \sum\limits_{i=1}^{j} \alpha(i) f(j-i), & \text{if } j > 0, \end{cases}$$

and

$$\alpha(j) = \frac{\tilde{c}_{s,S} g_R(j)}{1 - \tilde{c}_{s,S} g_R(0)}.$$

The next step is to relate the inventory position to the on-hand inventory level. The approximate steady-state distribution of IL^+ can be derived conditionally on IP^+

$$\tilde{\pi}^+_{IL}(j) = \begin{cases} \sum\limits_{i=j}^{S} \tilde{c}_{s,S} g_L(i-j) \tilde{\pi}^+_{IP}(i), & \text{if } j > 0, \\ \sum\limits_{i=j}^{S} \left(1 - \tilde{c}_{s,S} \mathcal{G}^0_L(i)\right) \tilde{\pi}^+_{IP}(i), & \text{if } j = 0. \end{cases} \quad (6.15)$$

Similarly, for IL^-

$$\tilde{\pi}^-_{IL}(j) = \begin{cases} \sum\limits_{i=j}^{S} \tilde{c}_{s,S} g_{L+R}(i-j) \tilde{\pi}^+_{IP}(i), & \text{if } j > 0, \\ \sum\limits_{i=j}^{S} \left(1 - \tilde{c}_{s,S} \mathcal{G}^0_{L+R}(i)\right) \tilde{\pi}^+_{IP}(i), & \text{if } j = 0. \end{cases} \quad (6.16)$$

These closed-form expressions approximate the steady-state behavior of the inventory system with a POUTP. Based on these equilibrium distributions, we can derive approximations for the performance measures. For instance, the approximate expected on-hand inventory level after order delivery equals

$$\begin{aligned}\tilde{IL}^+_{s,S} &= \sum_{i=0}^{S} i\tilde{\pi}^+_{IL}(i) = \sum_{i=1}^{S}\sum_{j=i}^{S} i\tilde{c}_{s,S}g_L(j-i)\tilde{\pi}^+_{IP}(j) \\ &= \sum_{j=0}^{S} \tilde{c}_{s,S}E\left[(j-D_L)^+\right]\tilde{\pi}^+_{IP}(j) = \sum_{j=0}^{S}\tilde{c}_{s,S}\mathcal{G}^1_L(j)\tilde{\pi}^+_{IP}(j),\end{aligned}$$

and before order delivery

$$\tilde{IL}^-_{s,S} = \sum_{i=0}^{S} i\tilde{\pi}^-_{IL}(i) = \sum_{j=0}^{S}\tilde{c}_{s,S}\mathcal{G}^1_{L+R}(j)\tilde{\pi}^+_{IP}(j). \tag{6.17}$$

Similar as in Section 5.3, the difference between the two average inventory levels expresses the increase of the on-hand inventory level due to an order. Hence, the approximate average order size equals $\tilde{IL}^+_{s,S} - \tilde{IL}^-_{s,S}$. Consequently, the fraction of demand lost is approximated by

$$\tilde{A}_{s,S} = 1 - \frac{\tilde{IL}^+_{s,S} - \tilde{IL}^-_{s,S}}{E[D_R]} = 1 - \frac{\sum_{j=0}^{S}\tilde{c}_{s,S}\left[\mathcal{G}^1_L(j) - \mathcal{G}^1_{L+R}(j)\right]\tilde{\pi}^+_{IP}(j)}{E[D_R]}. \tag{6.18}$$

Another important performance measure is the expected on-hand inventory level. Similar to Section 6.1, this expression depends on the type of demand distribution. For a (compound or pure) Poisson demand process, the approximate expected number of units on hand per unit time equals

$$\tilde{IL}_{s,S} = \frac{1}{R}\sum_{i=s+1}^{S}\sum_{j=1}^{i}\tilde{c}_{s,S}\left[g_L(i-j) - g_{L+R}(i-j)\right]H(j)\pi^+_{IP}(i), \tag{6.19}$$

whereas for a negative binomial demand distribution (or any other demand distribution)

$$\tilde{IL}_{s,S} = \frac{\tilde{IL}^+_{s,S} + \tilde{IL}^-_{s,S}}{2} = \frac{\sum_{j=0}^{S}\tilde{c}_{s,S}\left[\mathcal{G}^1_L(j) + \mathcal{G}^1_{L+R}(j)\right]\tilde{\pi}^+_{IP}(j)}{2}. \tag{6.20}$$

Based on these approximate expressions for the inventory levels, the expected total costs are approximated by

$$\begin{aligned}\tilde{C}_{s,S} &= K\sum_{i=0}^{s}\tilde{\pi}_{IL}^{-}(i) + h\tilde{IL}_{s,S} + pE[D]\tilde{A}_{s,S}\\ &= K\sum_{i=s+1}^{S}\big(1-\tilde{c}_{s,S}\mathcal{G}_{R}^{0}(i-s)\big)\tilde{\pi}_{IP}^{+}(i) + h\tilde{IL}_{s,S} + pE[D]\tilde{A}_{s,S}.\end{aligned}$$

The final step of our approximation procedure is to compute $\tilde{c}_{s,S}$. As in Section 5.3, we propose two procedures to set the value of $\tilde{c}_{s,S}$. Notice that there is not a unique value of $\tilde{c}_{s,S}$. When $\tilde{c}_{s,S}$ equals 1, the expressions derived in this section represent the expected total costs and fill rate for a backorder model (see also Section 5.3). However, for a lost-sales inventory system $\tilde{c}_{s,S}$ should be larger than one. Just consider Equation (6.15) and Equation (6.16), the probability for the on-hand inventory level to be strictly larger than zero is higher in a lost-sales system compared to a backorder system. Therefore, $\tilde{c}_{s,S} > 1$.

In the first approach to determine $\tilde{c}_{s,S}$, we use Little's formula and equate the approximate average inventory on order with the approximate average amount ordered per unit time multiplied by the lead time L. Similar to Section 5.3, we consider the inventory status during a shifted review period $[T+L, T+L+R)$, since all outstanding orders at time T are delivered at or before time $T+L$ and no orders are delivered during the shifted review period. Based on our approximations, the expected on-hand inventory at review instant in this period is approximated by

$$\sum_{i=s+1}^{S}\sum_{j=1}^{i} j\tilde{c}_{s,S}g_{L+R-r}(i-j)\tilde{\pi}_{IP}^{+}(i) = \sum_{i=s+1}^{S}\tilde{c}_{s,S}\mathcal{G}_{L+R-r}^{1}(i)\tilde{\pi}_{IP}^{+}(i), \tag{6.21}$$

whereas the approximate inventory position at the same review instant equals $\tilde{IP}_{s,S}^{+} = \sum_{i} i\tilde{\pi}_{IP}^{+}(i)$. Therefore, the average inventory on order during $[T+L+R-r, T+L+R)$ is approximated by

$$\sum_{i=s+1}^{S}\Big[i - \tilde{c}_{s,S}\mathcal{G}_{L+R-r}^{1}(i)\Big]\tilde{\pi}_{IP}^{+}(i). \tag{6.22}$$

As mentioned before, the average order size is expressed as $IL_{s,S}^{+} - IL_{s,S}^{-}$. Consequently, this amount should be subtracted from Equation (6.22) to represent the approximate inventory on order before order placement at review time $T+L+R-r$. Therefore, the approximation for the average inventory on order equals the time-weighted average of the two average inventory levels on order (before and after order placement), i.e.,

$$\begin{aligned}\tilde{O}_{s,S} &= \sum_{i=s+1}^{S}\left[i-\tilde{c}_{s,S}\mathcal{G}^1_{L+R-r}(i)\right]\tilde{\pi}^+_{IP}(i)-\frac{R-r}{R}\left[\tilde{IL}^+_{s,S}-\tilde{IL}^-_{s,S}\right],\\ &= \sum_{i=s+1}^{S}\left\{i-\tilde{c}_{s,S}\left[\mathcal{G}^1_{L+R-r}(i)+\frac{R-r}{R}\left(\mathcal{G}^1_{L}(i)-\mathcal{G}^1_{L+R}(i)\right)\right]\right\}\tilde{\pi}^+_{IP}(i).\end{aligned}$$

Notice that $\tilde{\pi}^+_{IP}$ also depends on $\tilde{c}_{s,S}$. Therefore, there is no explicit expression for $\tilde{c}_{s,S}$. With a bisection method we can find the value of $\tilde{c}_{s,S}$ such that

$$\tilde{O}_{s,S}=\frac{L}{R}\left[\tilde{IL}^+_{s,S}-\tilde{IL}^-_{s,S}\right], \tag{6.23}$$

based on Little's formula.

A second procedure to set the value of $\tilde{c}_{s,S}$ is based on the average order size. The average order size equals the expected increase in the on-hand inventory level at an order delivery $(\tilde{IL}^+_{s,S}-\tilde{IL}^-_{s,S})$, but also the expected increase in the inventory position at a review instant $(\tilde{IP}^+_{s,S}-\tilde{IP}^-_{s,S})$. In this second approach to determine the value of $\tilde{c}_{s,S}$, the following equality should be satisfied:

$$\sum_{i=s+1}^{S}\tilde{c}_{s,S}\left[\mathcal{G}^1_{L}(i)-\mathcal{G}^1_{L+R}(i)\right]\tilde{\pi}^+_{IP}(i)=\sum_{i=s+1}^{S}\left[i-\tilde{c}_{s,S}\mathcal{G}^1_{R}(i)\right]\tilde{\pi}^+_{IP}(i). \tag{6.24}$$

A bisection method could be used to find the value of $\tilde{c}_{s,S}$ which satisfies Equation (6.24).

To summarize, when a cost objective is considered in an inventory control system with POUTP we want to find the value of reorder level s and order-up-to level S such that the approximate expected total costs are minimized. Let us denote these values by $\tilde{s}$ and $\tilde{S}$, respectively. Hence, $\tilde{s}=s(\tilde{S})$ and

$$\tilde{S}=\underset{S\geq 0}{\operatorname{argmin}}\left\{\tilde{C}_{s(S),S}\right\}, \tag{6.25}$$

where

$$s(S)=\underset{0\leq s<S}{\operatorname{argmin}}\left\{\tilde{C}_{s,S}\right\}.$$

The value of $\tilde{c}_{s,S}$ can be based on either Equation (6.23) or Equation (6.24), which we denote by either subscript 1 or 2, respectively. Similarly for the service model $\tilde{s}=s(\tilde{S})$ and

$$\tilde{S}=\underset{S\geq 0}{\operatorname{argmin}}\left\{\tilde{C}_{s(S),S}\,\middle|\,1-\tilde{A}_{s(S),S}\geq\overline{\beta}\right\},$$

where

$$s(S) = \operatorname*{argmin}_{0 \le s < S} \left\{ \tilde{C}_{s,S} \middle| 1 - \tilde{A}_{s,S} \ge \overline{\beta} \right\}.$$

As explained before, when $\tilde{c}_{s,S} = 1$ the expressions correspond to a backorder model. The best values of the reorder level and order-up-to level in such a model are denoted s_{BO} and S_{BO}.

To illustrate the performance of our approximation procedure and the backorder model, we have computed the exact and approximate expected total costs and fill rate for different order-up-to levels and their corresponding optimal reorder levels. The same example is used as discussed in the previous section. The results are shown in Figure 6.3. The approximation procedure where $\tilde{c}_{s,S}$ is based on Equation (6.23) seems to approximate the actual costs and fill rate better compared to the procedure where $\tilde{c}_{s,S}$ is based on Equation (6.24). As in Section 5.3, the backorder model overestimates the expected costs. However, the fill rate is not underestimated for all values of S. When each of these approximation procedures is used to find near-optimal inventory control values, Table 6.7 provides a summary of the results. Surprisingly, the approximation procedure based on Equation (6.23) performs worse than based on Equation (6.24). Furthermore, the backorder model performs better than both approximation procedures. As will be shown in Section 6.4, this result is not a general conclusion.

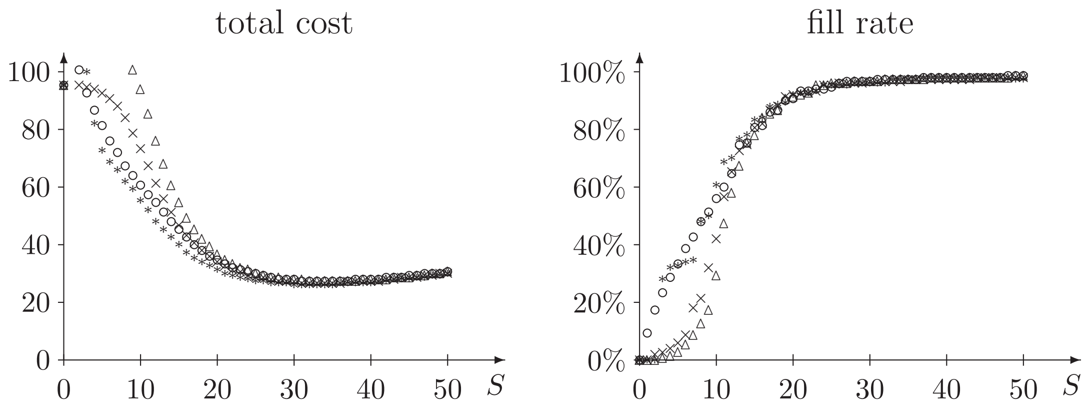


Figure 6.3: The expected total costs and fill rate for different values of S in the POUTP when the exact (circle), approximation (with Eq. (6.23): astrics, or Eq. (6.24): cross) or backorder model (triangle) is used where $L = 1.5R$, $\lambda = 5$, $\mu = 1$, $h = 1$, $p = 19$ and $K = 50$.

Up till now we have only considered inventory systems with a POUTP. As illustrated in Table 6.5, the same reorder level and order-up-to level could be used for the ROUTP. However, the value of maximum order quantity q should be specified as well. Similar to Section 5.3, we relate the order-up-to level to

	s	S	cost
exact	12	34	27.36
approximation - Eq. (6.23)	10	32	28.11
approximation - Eq. (6.24)	11	33	27.55
backorder model	12	34	27.36

Table 6.7: The values of the inventory control variables based on the approximation procedures for POUTPs.

the length of a replenishment cycle to determine this value. Let N denote the average number of review periods in between two consecutive orders. This number depends on the actual order size, which consists of $S-s$ units and the *undershoot* (i.e., the amount of inventory below reorder level s at a review instant). We assume that it is as likely to occur that the inventory level drops down to the reorder level at the start of a review period as at the end. Hence,

$$N = \frac{S-s}{E[D_R]} + \frac{Var[D_R] + E[D_R]^2}{2E[D_R]^2}, \tag{6.26}$$

since the average undershoot equals $(Var[D_R] + E[D_R]^2)/(2E[D_R])$ (see Tijms and Groenevelt [287], Johansen and Hill [134]).

It takes on average N review periods before a new order is placed and an extra L time units before it gets delivered. Therefore, the inventory position after ordering should cover the demand during this period. Hence, we propose to set the maximum order quantity q equal to $SRN/(L+RN)$ rounded to the nearest integer (rounded up in case of a tie). However, the maximum order size should at least be $S-s$, otherwise the policy corresponds to a FOSP. Consequently,

$$q = \max\left\{round\left(S\frac{RN}{L+RN}\right), S-s\right\}. \tag{6.27}$$

These expressions to set the value of q can be applied to the examples discussed in Table 6.7. The results are presented in Table 6.8. When the results in both tables are compared, we see almost no differences in costs when a ROUTP is applied. More numerical results are discussed in the next section.

6.4 Numerical results

The goal of this section is twofold. First, we compare the performance of the different replenishment policies. Second, the performance of the approximation procedures is investigated to find near-optimal values of reorder level s, order-up-to level S and maximum order quantity q. The performance of the policies is

	s	S	q	cost
exact	12	34	26	27.34
approximation - Eq. (6.23)	10	32	26	28.13
approximation - Eq. (6.24	11	33	25	27.56
backorder	12	34	26	27.34

Table 6.8: The values of the inventory control variables based on the approximation procedures for ROUTPs.

illustrated for a test bed similar to the one in Section 5.4. However, the penalty cost p is fixed to 19 and the fixed order cost K has values of 25, 50, and 100. Furthermore, $R = 1$ and L ranges from 0.5 to 3.5, holding cost h equals 1. The same demand distributions are used as presented in Table 5.5. So, pure and compound Poisson demand processes and negative binomial distributions.

Cost model

For the POUTP, we compute reorder level $\overline{s}$ and order-up-to level $\overline{S}$ that minimize the expected total costs as described in Section 6.2.2. The values of s and S that minimize the approximated costs $\tilde{C}_{s,S}$, where $\tilde{c}_{s,S}$ is based on Equation (6.23) or Equation (6.24), is denoted by $\tilde{s}_1$ and $\tilde{S}_1$, or by $\tilde{s}_2$ and $\tilde{S}_2$, respectively. The values based on the backorder model are denoted s_{BO} and S_{BO}. For the ROUTP, we compute the service level s^*, order-up-to level S^* and maximum order size q^* that minimize the average costs (see Section 6.2.3). We also investigate four other ROUTPs in which the reorder levels and order-up-to levels are based on the models described above (for the POUTP). Their maximum order sizes q (denoted $\overline{q}$, $\tilde{q}_1$, $\tilde{q}_2$, and q_{BO}, respectively) are set according to Equation (6.27). We also report the average costs C^* and fill rate β^* for an optimal replenishment policy, which is computed according to Section 6.2.1.

Table 6.11 to Table 6.16 provide the results for the pure Poisson, compound Poisson and negative binomial demand process, respectively. It shows the optimal costs C^*, the parameter values for each of the beforementioned replenishment policies and the corresponding cost increases compared to the costs of an optimal policy. We have also included the results obtained when the policy improvement algorithm of Hill and Johansen [118] is applied to find the replenishment policy. The approximation procedure to determine the values of the reorder level and fixed order quantity in the FOSP based on Johansen and Hill [134] is also included, which is denoted by s' and Q' respectively. Table 6.9 gives a summary of the results, where the average is taken over the relative cost increases for each of the policies compared to the optimal policy.

		[118]	$\hat{s},\hat{Q}$	$\bar{s},\bar{S}$	$\tilde{s}_1,\tilde{S}_1$	$\tilde{s}_2,\tilde{S}_2$	$s_{\text{BO}},S_{\text{BO}}$	s^*,S^*,q^*	$\bar{s},\bar{S},\bar{q}$	$\tilde{s}_1,\tilde{S}_1,\tilde{q}_1$	$\tilde{s}_2,\tilde{S}_2,\tilde{q}_2$	$s_{\text{BO}},S_{\text{BO}},q_{\text{BO}}$
pure	average	0.4%	0.3%	0.2%	3.7%	0.6%	0.4%	0.1%	0.2%	3.9%	0.5%	0.3%
Poisson	st.dev.	1.7%	0.4%	0.4%	3.1%	0.5%	0.5%	0.3%	0.4%	3.4%	0.5%	0.5%
	max	9.7%	1.5%	1.5%	10.8%	2.1%	2.3%	1.4%	1.5%	12.1%	1.9%	2.2%
compound	average	0.4%	0.4%	0.2%	1.9%	0.6%	0.5%	0.1%	0.2%	2.1%	0.4%	0.3%
Poisson	st.dev.	1.3%	0.4%	0.3%	1.7%	0.4%	0.5%	0.3%	0.3%	1.9%	0.4%	0.5%
	max	7.5%	1.2%	1.3%	6.8%	1.7%	2.4%	1.1%	1.2%	7.5%	1.5%	2.2%
negative	average	0.4%	0.3%	0.2%	2.2%	0.5%	0.5%	0.1%	0.2%	2.4%	0.4%	0.3%
binomial	st.dev.	1.5%	0.4%	0.3%	2.2%	0.4%	0.5%	0.3%	0.3%	2.5%	0.4%	0.5%
	max	8.4%	1.2%	1.3%	8.8%	1.7%	2.1%	1.2%	1.2%	11.6%	1.8%	2.0%

Table 6.9: A summary of the results for the cost model: the average, standard deviation and maximum cost increase of each policy compared to the optimal policy.

Similar observations can be made as in Section 5.4. First, the order-up-to levels $\tilde{S}_1$ are always less than or equal to $\overline{S}$. The backorder model results in order-up-to levels higher than or equal to $\overline{S}$. Hence, $\tilde{S}_1 \leq \overline{S} \leq S_{BO}$, but also $\tilde{S}_1 \leq \tilde{S}_2 \leq S_{BO}$. The numerical experiments show that the best ROUTP results in policies with almost the same average costs as the optimal policy. We can conclude the same for ROUTPs specified by $\overline{s}$, $\overline{S}$ and $\overline{q}$. Surprisingly, adding this extra restriction on the maximum order size (q) does not show much effect. This is mainly because the POUTP performs close to optimal, and the ROUTP cannot improve the results much. When we compare these order-up-to policies with fixed order size policies, we see that FOSPs perform similarly well. However, the approximation procedure of Johansen and Hill [134] to determine the values of the inventory control variables in the FOSP performs poorly. The approximation procedures for the order-up-to policies (as described in Section 6.3) result in near-optimal values for the order sizes. It is clear that the approximation procedure based on Equation (6.24) and the backorder model are superior compared to the approximation procedure based on Equation (6.23). The results are not decisive whether the approximation procedure or the backorder model performs better. The cost increases in the ROUTPs based on the approximation procedure are on average around 0.4% compared to the optimal costs. Similar to Chapter 5, the cost increases are more significant when the value of q is underestimated compared to the optimal value q^*. This is also the reason that the ROUTP based on the approximation procedure with Equation (6.23) performs worse than the POUTP based on the same procedure to set the values of the inventory control variables.

In conclusion, order-up-to policies and fixed order size policies perform close to optimal (around 0.3% deviation). The approximation procedure derived in Section 6.3 and the backorder model result in near-optimal order sizes, with a cost increase of around 0.4% compared to the optimal policy. Therefore, we conclude that (restricted) order-up-to policies should be applied in practice.

Service model

Similar to Section 5.4, we restrict to a similar test bed for the service model due to the higher complexity involved to solve the service model (see Section 6.2.1). We set $h = 1$, $p = 0$, $K = 50$, and the minimal fill rate $\overline{\beta}$ equals 85%, 90%, and 95%. The objective is to minimize the total expected costs (consisting of holding and order costs) subject to the service level requirement as formulated in Section 6.2.1. For each problem instance, we compute the optimal policy and the optimal control values for each alternative replenishment policy (FOSP, POUTP, ROUTP) based on the exact and approximation models. Similar notation is used as for the cost model. We have also included the inventory control values based on the approximation procedure by Tijms and Groenevelt [287] for the POUTP (denoted s', S'). The resulting expected costs for each inventory control policy is presented in Table 6.17 to Table 6.20 for Poisson and negative binomial distributed demand. A summary is provided in Table 6.10, where the average is taken over the relative increase of the expected average costs for each of the policies compared to an optimal policy. This is only possible for the instances where the service level requirement is met. Therefore, we also included the number of instances that the requirement is not satisfied. Note that there are 27 instances for pure Poisson demand, 9 instances for compound Poisson and 27 instances for the negative binomial distributed demand.

Based on these results, we conclude that the cost increases for the FOSP, POUTP and ROUTP compared to the optimal replenishment policy are on average 0.70%, 0.86%, and 0.45%, respectively. To find near-optimal values of the inventory control variables we recommend to use the backorder model for the ROUTP where q is derived according to Equation (6.27). This approximation procedure results in a cost increase of around 1.5%. However, the service level constraint is guaranteed in (almost) all test instances.

6.5 Concluding remarks

In the previous chapter no fixed order cost was assumed to find optimal and near-optimal replenishment policies for a periodic review inventory system where excess demand is lost. In this chapter we extended these results to the case with fixed order cost. A general cost and service models is developed to find an optimal policy and compute different performance measures of interest such as the average costs and fill rate. The computational effort to determine an optimal replenishment policy is however large. Therefore, we proposed several alternative replenishment

		$\bar{s}, \bar{S}$	$\hat{s}, \hat{Q}$	s^*, S^*, q^*	$\bar{s}, \bar{S}, \bar{q}$	s', S'	$\tilde{s}_1, \tilde{S}_1$	$\tilde{s}_1, \tilde{S}_1, \tilde{q}_1$	$\tilde{s}_2, \tilde{S}_2$	$\tilde{s}_2, \tilde{S}_2, \tilde{q}_2$	$s_{\mathrm{BO}}, S_{\mathrm{BO}}$	$s_{\mathrm{BO}}, S_{\mathrm{BO}}, q_{\mathrm{BO}}$
pure	average	0.95%	0.77%	0.61%	1.70%	1.44%	3.09%	2.37%	3.09%	3.26%	2.81%	1.57%
Poisson	frequency	-	-	-	20	13	22	22	18	22	-	1
compound	average	0.7%	0.44%	0.09%	0.13%	0.86%	-	-	2.04%	-	2.62%	1.24%
Poisson	frequency	-	-	-	6	5	9	9	6	9	-	-
negative	average	0.82%	0.72%	0.41%	1.05%	1.10%	1.05%	0.97%	2.08%	1.11%	3.09%	1.61%
binomial	frequency	-	-	-	22	19	21	26	16	24	-	-

Table 6.10: A summary of the results for the service model: the average increase in average inventory level of each policy compared to the optimal policy when the service constraint is met, including the frequency that the service level is not met.

policies based on fixed order sizes and order-up-to levels. In particular, we proposed a new replenishment policy in which the maximum order quantities are restricted to an upper limit. Such policies result in near-optimal costs (0.1% cost increase in the cost model and 0.45% in the service model). Therefore, such restricted order-up-to policies (ROUTPs) are recommended to be used in practical settings. We also developed approximation procedures to set near-optimal values of the inventory control variables for the ROUTP, which result in an average cost increase of 0.5% from the optimal costs in the cost model. However, in some practical settings (restricted) order-up-to policies are not preferred. Consequently, the replenishment policy should satisfy the restrictions in such circumstances. This is the topic of the next chapter.

λ, μ	L	K	C^*	β^*	[118]	$\bar{s}, \bar{S}$	$C_{\bar{s},\bar{S}}$	$\hat{s}, \hat{Q}$	$C_{\hat{s},\hat{Q}}$	s^*, S^*, q^*	$C_{s*,S*,q*}$	$C_{\bar{s},\bar{S},\bar{q}}$
2,1	0.5	25	12.39	97.6%	0.00%	3,13	0.05%	3,11	0.23%	3,13,12	0.00%	0.00%
2,1	0.5	50	16.11	96.3%	0.00%	2,16	0.00%	2,15	0.01%	2,16,16	0.00%	0.01%
2,1	0.5	100	21.56	97.2%	0.00%	2,22	0.03%	2,20	0.05%	2,22,21	0.00%	0.00%
2,1	1.5	25	12.98	96.1%	0.00%	5,15	0.06%	5,11	0.14%	5,15,12	0.00%	0.00%
2,1	1.5	50	16.63	94.9%	0.00%	4,18	0.09%	4,15	0.05%	4,19,16	0.00%	0.10%
2,1	1.5	100	21.95	93.9%	0.00%	3,23	0.05%	3,21	0.00%	3,24,21	0.00%	0.12%
2,1	2.5	25	13.45	95.0%	0.00%	7,17	0.09%	7,11	0.15%	7,18,12	0.01%	0.07%
2,1	2.5	50	17.03	94.0%	0.00%	6,21	0.11%	6,16	0.07%	6,21,16	0.00%	0.00%
2,1	2.5	100	22.22	93.2%	0.00%	5,25	0.07%	5,21	0.01%	5,26,21	0.01%	0.15%
2,1	3.5	25	13.85	94.1%	0.00%	9,20	0.10%	9,12	0.16%	9,20,13	0.02%	0.02%
2,1	3.5	50	17.35	93.2%	0.00%	8,23	0.08%	8,16	0.05%	8,23,16	0.00%	0.00%
2,1	3.5	100	22.45	92.7%	0.00%	7,27	0.07%	7,21	0.01%	7,28,22	0.01%	0.17%
5,1	0.5	25	20.43	97.7%	0.00%	7,22	0.04%	7,18	0.32%	7,22,20	0.00%	0.01%
5,1	0.5	50	26.41	98.3%	0.00%	7,28	0.02%	7,24	0.19%	7,28,26	0.00%	0.01%
5,1	0.5	100	35.00	98.0%	0.00%	6,36	0.03%	6,33	0.06%	6,37,34	0.00%	0.04%
5,1	1.5	25	21.50	97.7%	0.00%	13,28	0.05%	13,18	0.34%	13,28,20	0.00%	0.00%
5,1	1.5	50	27.34	97.4%	0.00%	12,34	0.05%	12,24	0.15%	12,34,26	0.00%	0.00%
5,1	1.5	100	35.81	97.2%	0.00%	11,42	0.04%	11,33	0.06%	11,42,35	0.00%	0.01%
5,1	2.5	25	22.34	96.9%	0.00%	18,34	0.38%	18,19	0.39%	18,34,19	0.19%	0.29%
5,1	2.5	50	28.15	96.6%	0.00%	17,39	0.06%	17,25	0.12%	17,40,26	0.01%	0.03%
5,1	2.5	100	36.48	96.5%	0.00%	16,47	0.05%	16,34	0.05%	16,48,35	0.00%	0.04%
5,1	3.5	25	22.90	97.1%	2.01%	24,39	1.08%	24,17	1.22%	24,39,19	1.04%	1.06%
5,1	3.5	50	28.78	96.9%	0.00%	23,45	0.18%	23,24	0.11%	23,47,24	0.11%	0.18%
5,1	3.5	100	37.05	96.0%	0.00%	21,53	0.04%	21,34	0.05%	21,53,36	0.00%	0.01%
10,1	0.5	25	30.18	98.7%	0.00%	15,33	0.07%	15,24	0.51%	15,33,26	0.00%	0.01%
10,1	0.5	50	38.67	98.6%	0.00%	14,42	0.05%	14,34	0.32%	14,43,37	0.00%	0.04%
10,1	0.5	100	50.91	98.5%	0.00%	13,55	0.03%	13,47	0.14%	13,55,50	0.00%	0.00%
10,1	1.5	25	31.61	98.4%	0.00%	26,44	0.55%	25,26	0.88%	26,45,27	0.33%	0.37%
10,1	1.5	50	40.07	98.4%	0.00%	25,54	0.08%	25,34	0.25%	25,54,36	0.00%	0.01%
10,1	1.5	100	52.16	97.9%	0.00%	23,66	0.04%	23,47	0.11%	23,66,51	0.00%	0.00%
10,1	2.5	25	32.69	98.3%	2.43%	37,54	1.04%	36,23	1.02%	36,55,24	0.85%	0.96%
10,1	2.5	50	41.09	97.9%	0.00%	35,64	0.54%	35,36	0.47%	35,66,36	0.23%	0.39%
10,1	2.5	100	53.19	97.9%	0.00%	34,77	0.02%	34,47	0.10%	34,77,53	0.00%	0.01%
10,1	3.5	25	33.48	97.9%	9.72%	47,64	1.50%	47,24	1.47%	47,66,24	1.40%	1.53%
10,1	3.5	50	41.87	97.9%	1.90%	46,74	1.05%	45,34	1.10%	46,75,35	0.94%	0.98%
10,1	3.5	100	54.07	97.5%	0.00%	44,88	0.14%	44,51	0.42%	44,89,45	0.03%	0.20%

Table 6.11: The results for the exact cost model when the demand follows a pure Poisson process.

λ, μ	L	K	s', Q'	$C_{s',Q'}$	$\tilde{s_1}, \tilde{S_1}$	$C_{\tilde{s_1},\tilde{S_1}}$	$\tilde{s_2}, \tilde{S_2}$	$C_{\tilde{s_2},\tilde{S_2}}$	s_{BO}, S_{BO}	$C_{s_{BO},S_{BO}}$	$C_{\tilde{s_1},\tilde{S_1},\tilde{q_1}}$	$C_{\tilde{s_2},\tilde{S_2},\tilde{q_2}}$	$C_{s_{BO},S_{BO},q_{BO}}$
2,1	0.5	25	4,11	4.04%	2,12	0.86%	2,12	0.86%	3,13	0.05%	0.96%	0.96%	0.00%
2,1	0.5	50	3,15	1.25%	2,16	0.00%	2,16	0.00%	2,17	0.12%	0.01%	0.01%	0.03%
2,1	0.5	100	2,21	0.05%	1,21	0.16%	1,21	0.16%	2,22	0.03%	0.24%	0.24%	0.00%
2,1	1.5	25	7,11	6.90%	4,14	1.76%	4,15	1.51%	5,16	0.34%	1.94%	1.42%	0.19%
2,1	1.5	50	6,15	3.39%	3,18	1.88%	4,19	0.14%	4,19	0.14%	2.00%	0.00%	0.00%
2,1	1.5	100	2,22	1.32%	2,22	1.41%	2,23	1.35%	4,25	0.25%	1.57%	1.34%	0.12%
2,1	2.5	25	11,11	16.77%	6,16	2.07%	7,18	0.17%	7,18	0.17%	2.34%	0.05%	0.05%
2,1	2.5	50	10,15	10.82%	4,19	4.88%	6,21	0.11%	6,22	0.42%	5.27%	0.00%	0.18%
2,1	2.5	100	3,23	2.71%	2,2	5.68%	4,26	1.16%	5,27	0.36%	6.20%	0.94%	0.07%
2,1	3.5	25	13,11	14.25%	7,18	5.78%	9,20	0.10%	9,21	0.56%	6.00%	0.02%	0.40%
2,1	3.5	50	13,15	14.29%	5,19	8.95%	8,23	0.08%	8,24	0.33%	10.02%	0.00%	0.12%
2,1	3.5	100	4,23	4.00%	5,2	5.30%	6,28	0.92%	7,29	0.33%	6.68%	0.71%	0.06%
5,1	0.5	25	11,17	10.19%	7,22	0.04%	7,22	0.04%	7,22	0.04%	0.01%	0.01%	0.01%
5,1	0.5	50	10,24	6.62%	6,27	0.27%	6,27	0.27%	7,28	0.02%	0.33%	0.33%	0.01%
5,1	0.5	100	10,32	6.57%	5,36	0.30%	5,36	0.30%	6,37	0.04%	0.29%	0.29%	0.00%
5,1	1.5	25	18,17	14.04%	12,27	0.98%	12,28	0.82%	13,29	0.13%	1.01%	0.75%	0.05%
5,1	1.5	50	18,23	13.28%	10,32	2.81%	11,33	0.77%	12,34	0.05%	2.89%	0.79%	0.00%
5,1	1.5	100	17,33	9.71%	9,40	1.77%	10,42	0.49%	11,43	0.09%	1.87%	0.43%	0.01%
5,1	2.5	25	25,17	18.57%	16,32	4.35%	18,34	0.38%	19,35	0.66%	4.47%	0.34%	0.61%
5,1	2.5	50	24,24	14.02%	15,37	3.12%	17,40	0.08%	18,40	0.08%	3.26%	0.02%	0.04%
5,1	2.5	100	24,33	13.10%	1,44	5.66%	15,48	0.56%	16,49	0.15%	6.36%	0.48%	0.05%
5,1	3.5	25	32,17	24.24%	21,37	4.25%	23,40	1.42%	24,40	1.23%	4.35%	1.49%	1.22%
5,1	3.5	50	31,24	18.28%	19,41	6.64%	22,45	0.30%	23,46	0.20%	7.44%	0.27%	0.15%
5,1	3.5	100	30,33	14.10%	1,46	10.79%	21,53	0.04%	22,54	0.10%	12.14%	0.01%	0.06%
10,1	0.5	25	21,24	12.84%	14,32	0.69%	14,32	0.69%	15,33	0.07%	0.63%	0.63%	0.01%
10,1	0.5	50	21,33	11.83%	13,41	0.50%	13,41	0.50%	14,43	0.05%	0.50%	0.50%	0.01%
10,1	0.5	100	20,46	8.73%	12,54	0.21%	12,54	0.21%	13,55	0.03%	0.19%	0.19%	0.00%
10,1	1.5	25	35,24	20.21%	23,42	4.81%	25,44	0.93%	26,46	0.68%	4.04%	0.63%	0.64%
10,1	1.5	50	34,33	15.43%	22,51	2.76%	24,53	0.28%	25,54	0.08%	2.72%	0.21%	0.03%
10,1	1.5	100	33,46	11.84%	21,63	1.37%	22,65	0.43%	23,67	0.07%	1.43%	0.41%	0.01%
10,1	2.5	25	47,25	22.67%	3,53	5.62%	36,56	1.43%	37,57	1.41%	5.41%	1.21%	1.27%
10,1	2.5	50	47,33	19.79%	3,61	6.14%	35,65	0.55%	36,66	0.65%	6.12%	0.44%	0.62%
10,1	2.5	100	46,46	15.30%	9,72	4.78%	33,77	0.19%	34,78	0.05%	5.13%	0.17%	0.02%
10,1	3.5	25	60,24	28.39%	4,62	8.99%	47,68	2.06%	48,69	2.34%	8.97%	1.90%	2.24%
10,1	3.5	50	59,34	22.46%	4,70	7.34%	45,76	1.35%	46,77	1.33%	7.36%	1.30%	1.24%
10,1	3.5	100	57,37	16.83%	7,80	9.55%	43,88	0.41%	44,89	0.17%	11.51%	0.42%	0.13%

Table 6.12: The results for the approximate cost model when the demand follows a pure Poisson process.

λ, μ	L	K	C^*	β^*	[118]	$\bar{s}, \bar{S}$	$C_{\bar{s},\bar{S}}$	$\hat{s}, \hat{Q}$	$C_{\hat{s},\hat{Q}}$	s^*, S^*, q^*	$C_{s*,S*,q*}$	$C_{\bar{s},\bar{S},\bar{q}}$
1,2	0.5	25	14.83	93.8%	0.00%	4,14	0.02%	3,12	0.37%	4,14,13	0.00%	0.00%
1,2	0.5	50	18.19	93.4%	0.00%	3,17	0.00%	3,16	0.14%	3,17,17	0.00%	0.02%
1,2	0.5	100	23.19	93.1%	0.00%	2,22	0.00%	2,21	0.03%	2,22,22	0.00%	0.02%
1,2	1.5	25	15.75	91.7%	0.00%	6,16	0.15%	6,12	0.32%	6,17,14	0.03%	0.13%
1,2	1.5	50	18.95	91.6%	0.00%	5,19	0.10%	5,16	0.12%	5,20,17	0.01%	0.12%
1,2	1.5	100	23.76	89.5%	0.00%	3,23	0.05%	3,22	0.02%	3,24,22	0.00%	0.29%
1,2	2.5	25	16.44	90.1%	0.00%	8,19	0.23%	8,13	0.31%	8,19,14	0.02%	0.09%
1,2	2.5	50	19.53	90.2%	0.00%	7,21	0.19%	7,16	0.13%	7,22,18	0.01%	0.20%
1,2	2.5	100	24.18	88.6%	0.00%	5,25	0.11%	5,22	0.02%	5,26,22	0.00%	0.33%
1,2	3.5	25	17.01	88.9%	0.00%	10,21	0.23%	10,13	0.25%	10,21,15	0.03%	0.16%
1,2	3.5	50	20.01	89.0%	0.00%	9,24	0.23%	8,17	0.11%	9,24,18	0.01%	0.03%
1,2	3.5	100	24.53	87.8%	0.00%	7,27	0.16%	7,22	0.03%	7,28,23	0.01%	0.37%
2.5,2	0.5	25	24.09	96.1%	0.00%	9,25	0.07%	9,19	0.50%	9,25,22	0.01%	0.01%
2.5,2	0.5	50	29.59	96.1%	0.00%	8,30	0.03%	8,25	0.24%	8,30,28	0.00%	0.01%
2.5,2	0.5	100	37.72	96.2%	0.00%	7,38	0.02%	7,34	0.10%	7,38,36	0.00%	0.01%
2.5,2	1.5	25	26.03	95.2%	0.00%	15,31	0.11%	15,20	0.49%	15,31,24	0.01%	0.04%
2.5,2	1.5	50	31.30	95.4%	0.00%	14,36	0.11%	14,26	0.26%	14,37,29	0.02%	0.04%
2.5,2	1.5	100	39.15	94.8%	0.00%	12,43	0.09%	12,35	0.09%	12,44,37	0.00%	0.08%
2.5,2	2.5	25	27.42	94.7%	0.08%	21,37	0.39%	20,21	0.68%	21,37,24	0.32%	0.34%
2.5,2	2.5	50	32.62	94.2%	0.00%	19,42	0.12%	19,27	0.23%	19,43,29	0.02%	0.05%
2.5,2	2.5	100	40.26	93.8%	0.00%	17,49	0.11%	17,35	0.09%	17,50,38	0.01%	0.05%
2.5,2	3.5	25	28.36	93.9%	1.57%	26,41	0.97%	26,18	1.15%	26,42,27	0.93%	1.05%
2.5,2	3.5	50	33.69	94.1%	0.00%	24,48	0.25%	25,26	0.24%	24,49,30	0.11%	0.17%
2.5,2	3.5	100	41.18	92.9%	0.00%	22,54	0.11%	22,36	0.08%	22,55,38	0.01%	0.09%
5,2	0.5	25	35.13	97.7%	0.00%	18,38	0.06%	18,27	0.86%	18,39,33	0.01%	0.02%
5,2	0.5	50	43.07	97.4%	0.00%	16,46	0.05%	16,36	0.39%	16,46,41	0.00%	0.01%
5,2	0.5	100	54.70	97.6%	0.00%	15,57	0.04%	14,49	0.17%	15,58,53	0.00%	0.02%
5,2	1.5	25	38.01	97.2%	0.02%	30,51	0.37%	28,29	0.93%	29,51,30	0.24%	0.28%
5,2	1.5	50	45.81	97.0%	0.00%	28,58	0.08%	28,37	0.41%	28,59,43	0.01%	0.03%
5,2	1.5	100	57.11	96.9%	0.00%	26,69	0.07%	26,49	0.18%	26,70,54	0.01%	0.02%
5,2	2.5	25	39.93	96.7%	2.53%	41,61	0.96%	40,26	1.21%	41,62,29	0.93%	0.95%
5,2	2.5	50	47.83	96.5%	0.01%	39,70	0.39%	38,39	0.43%	38,71,39	0.28%	0.32%
5,2	2.5	100	59.02	96.0%	0.00%	37,81	0.06%	36,51	0.16%	36,81,55	0.01%	0.02%
5,2	3.5	25	41.40	96.2%	7.51%	51,71	1.26%	51,26	1.25%	51,72,31	1.11%	1.23%
5,2	3.5	50	49.20	96.1%	1.80%	49,79	1.02%	49,35	1.16%	49,79,50	1.00%	1.05%
5,2	3.5	100	60.59	95.7%	0.00%	47,93	0.17%	47,48	0.15%	47,94,55	0.11%	0.11%

Table 6.13: The results for the exact cost model when the demand follows a compound Poisson process.

λ, μ	L	K	s', Q'	$C_{s',Q'}$	$\tilde{s_1}, \tilde{S_1}$	$C_{\tilde{s_1},\tilde{S_1}}$	$\tilde{s_2}, \tilde{S_2}$	$C_{\tilde{s_2},\tilde{S_2}}$	s_{BO}, S_{BO}	$C_{s_{BO},S_{BO}}$	$C_{\tilde{s_1},\tilde{S_1},\tilde{q_1}}$	$C_{\tilde{s_2},\tilde{S_2},\tilde{q_2}}$	$C_{s_{BO},S_{BO},q_{BO}}$
1,2	0.5	25	8,11	14.60%	3,14	0.18%	3,14	0.18%	4,14	0.02%	0.15%	0.15%	0.00%
1,2	0.5	50	0,19	5.72%	2,17	0.31%	2,17	0.31%	3,18	0.11%	0.34%	0.34%	0.04%
1,2	0.5	100	0,23	1.67%	1,22	0.25%	1,22	0.25%	2,23	0.09%	0.28%	0.28%	0.03%
1,2	1.5	25	13,11	26.29%	5,16	0.49%	5,17	0.67%	6,17	0.24%	0.47%	0.39%	0.03%
1,2	1.5	50	0,20	9.76%	4,19	0.29%	4,20	0.45%	5,21	0.52%	0.33%	0.22%	0.20%
1,2	1.5	100	0,24	3.61%	2,23	0.65%	3,24	0.09%	3,25	0.28%	0.75%	0.06%	0.02%
1,2	2.5	25	0,20	23.00%	7,18	0.63%	7,19	0.69%	8,20	0.57%	0.65%	0.39%	0.17%
1,2	2.5	50	0,21	13.02%	5,21	1.16%	6,22	0.42%	7,23	0.51%	1.05%	0.14%	0.12%
1,2	2.5	100	0,24	5.41%	3,23	1.34%	4,26	0.56%	5,28	0.74%	1.76%	0.34%	0.12%
1,2	3.5	25	0,20	26.06%	8,20	1.76%	9,22	0.95%	10,23	0.99%	1.94%	0.42%	0.46%
1,2	3.5	50	0,22	15.60%	6,22	2.17%	8,25	0.68%	9,26	0.91%	2.54%	0.12%	0.35%
1,2	3.5	100	0,25	6.97%	4,20	4.48%	5,29	1.45%	7,30	0.78%	6.55%	0.82%	0.23%
2.5,2	0.5	25	17,17	19.09%	9,24	0.08%	9,24	0.08%	9,25	0.07%	0.10%	0.10%	0.01%
2.5,2	0.5	50	16,24	15.67%	7,29	0.46%	7,29	0.46%	8,31	0.11%	0.54%	0.54%	0.02%
2.5,2	0.5	100	15,33	12.61%	6,38	0.13%	6,38	0.13%	7,39	0.07%	0.10%	0.10%	0.01%
2.5,2	1.5	25	26,18	25.71%	14,30	0.67%	14,31	0.60%	15,32	0.20%	0.65%	0.48%	0.04%
2.5,2	1.5	50	25,24	22.12%	12,35	0.93%	13,36	0.25%	14,38	0.30%	0.95%	0.16%	0.11%
2.5,2	1.5	100	24,33	18.50%	10,42	0.91%	11,44	0.33%	12,45	0.18%	0.94%	0.21%	0.02%
2.5,2	2.5	25	34,18	30.17%	18,35	2.47%	20,38	0.62%	21,39	0.68%	2.57%	0.46%	0.51%
2.5,2	2.5	50	33,25	26.56%	16,40	2.31%	18,43	0.50%	19,44	0.32%	2.37%	0.31%	0.12%
2.5,2	2.5	100	32,34	22.77%	13,46	2.82%	16,50	0.40%	17,52	0.39%	3.23%	0.23%	0.13%
2.5,2	3.5	25	42,19	36.90%	23,41	2.79%	26,44	1.35%	27,45	1.75%	2.88%	1.52%	1.86%
2.5,2	3.5	50	41,25	31.30%	20,45	3.77%	24,49	0.27%	25,50	0.43%	4.00%	0.12%	0.26%
2.5,2	3.5	100	3,51	26.24%	16,49	5.22%	21,56	0.44%	23,58	0.40%	6.30%	0.24%	0.20%
5,2	0.5	25	29,24	20.04%	16,37	1.01%	16,37	1.01%	18,39	0.09%	1.04%	1.04%	0.01%
5,2	0.5	50	28,34	17.19%	15,45	0.38%	15,45	0.38%	16,47	0.09%	0.36%	0.36%	0.01%
5,2	0.5	100	27,47	14.30%	13,56	0.36%	13,56	0.36%	15,58	0.05%	0.37%	0.37%	0.00%
5,2	1.5	25	45,25	26.80%	27,48	1.70%	28,50	0.76%	30,52	0.44%	1.70%	0.69%	0.34%
5,2	1.5	50	44,34	23.22%	25,55	1.52%	26,58	0.57%	28,60	0.16%	1.61%	0.49%	0.05%
5,2	1.5	100	43,47	19.90%	22,66	1.64%	24,69	0.37%	26,71	0.13%	1.70%	0.30%	0.03%
5,2	2.5	25	61,25	35.59%	36,59	3.76%	40,63	1.19%	41,64	1.29%	3.82%	1.20%	1.30%
5,2	2.5	50	59,35	28.51%	34,66	3.17%	38,70	0.41%	39,72	0.48%	3.29%	0.34%	0.39%
5,2	2.5	100	57,37	24.18%	31,76	2.89%	35,81	0.25%	37,83	0.14%	3.08%	0.17%	0.04%
5,2	3.5	25	75,19	40.33%	46,69	4.05%	51,75	1.73%	53,77	2.38%	3.92%	1.51%	2.18%
5,2	3.5	50	72,22	35.05%	43,76	4.39%	49,82	1.20%	50,84	1.49%	4.43%	1.25%	1.55%
5,2	3.5	100	13,81	47.22%	38,84	6.80%	46,93	0.22%	47,95	0.25%	7.53%	0.15%	0.15%

Table 6.14: The results for the approximate cost model when the demand follows a compound Poisson process.

w,u	L	K	C^*	β^*	[118]	$\bar{s},\bar{S}$	$C_{\bar{s},\bar{S}}$	$\hat{s},\hat{Q}$	$C_{\hat{s},\hat{Q}}$	s^*,S^*,q^*	$C_{s*,S*,q*}$	$C_{\bar{s},\bar{S},\bar{q}}$
2,0.5	0.5	25	13.71	94.4%	0.00%	3,13	0.02%	3,12	0.23%	3,13,13	0.02%	0.05%
2,0.5	0.5	50	17.25	93.5%	0.00%	2,17	0.03%	2,16	0.06%	2,17,16	0.00%	0.00%
2,0.5	0.5	100	22.42	95.0%	0.00%	2,22	0.01%	2,21	0.02%	2,22,21	0.00%	0.00%
2,0.5	1.5	25	14.54	92.5%	0.00%	5,16	0.16%	5,12	0.17%	5,16,13	0.00%	0.00%
2,0.5	1.5	50	17.92	91.8%	0.00%	4,19	0.11%	4,16	0.05%	4,19,17	0.00%	0.03%
2,0.5	1.5	100	22.92	91.5%	0.00%	3,23	0.06%	3,21	0.03%	3,24,22	0.00%	0.21%
2,0.5	2.5	25	15.16	93.2%	0.00%	8,18	0.16%	7,12	0.22%	8,19,13	0.03%	0.19%
2,0.5	2.5	50	18.43	90.8%	0.02%	6,21	0.15%	6,16	0.06%	6,21,17	0.01%	0.07%
2,0.5	2.5	100	23.28	90.7%	0.00%	5,25	0.11%	5,22	0.03%	5,26,22	0.00%	0.25%
2,0.5	3.5	25	15.65	92.0%	0.00%	10,21	0.19%	10,12	0.24%	10,21,14	0.02%	0.03%
2,0.5	3.5	50	18.84	89.8%	0.03%	8,23	0.17%	8,16	0.09%	8,24,17	0.02%	0.12%
2,0.5	3.5	100	23.59	90.0%	0.00%	7,27	0.13%	7,22	0.03%	7,28,22	0.00%	0.28%
2,2/7	0.5	25	25.07	96.0%	0.00%	10,25	0.05%	9,20	0.53%	10,26,23	0.01%	0.04%
2,2/7	0.5	50	30.40	95.4%	0.00%	8,30	0.04%	8,26	0.25%	8,31,28	0.00%	0.06%
2,2/7	0.5	100	38.36	95.6%	0.00%	7,38	0.02%	7,34	0.11%	7,39,36	0.00%	0.03%
2,2/7	1.5	25	27.10	95.0%	0.00%	16,32	0.16%	15,20	0.52%	16,32,25	0.04%	0.06%
2,2/7	1.5	50	32.21	94.5%	0.00%	14,37	0.13%	14,26	0.26%	14,37,30	0.01%	0.01%
2,2/7	1.5	100	39.89	94.1%	0.00%	12,44	0.11%	12,35	0.10%	12,44,38	0.01%	0.03%
2,2/7	2.5	25	28.56	93.7%	0.07%	21,37	0.38%	20,21	0.72%	21,38,25	0.28%	0.35%
2,2/7	2.5	50	33.60	93.3%	0.03%	19,42	0.17%	19,27	0.23%	19,43,31	0.01%	0.13%
2,2/7	2.5	100	41.06	93.0%	0.00%	17,49	0.14%	17,36	0.09%	17,50,38	0.01%	0.08%
2,2/7	3.5	25	29.52	93.5%	1.49%	26,42	0.99%	26,19	1.21%	26,42,27	0.92%	1.11%
2,2/7	3.5	50	34.72	93.1%	0.00%	24,49	0.30%	25,27	0.27%	24,49,31	0.12%	0.13%
2,2/7	3.5	100	42.03	92.1%	0.02%	22,55	0.14%	22,36	0.10%	22,56,39	0.01%	0.05%
10,0.5	0.5	25	32.93	98.3%	0.00%	17,37	0.07%	16,26	0.71%	17,37,31	0.01%	0.01%
10,0.5	0.5	50	41.08	97.9%	0.00%	15,44	0.05%	15,35	0.35%	15,45,40	0.00%	0.04%
10,0.5	0.5	100	52.97	98.0%	0.00%	14,56	0.03%	14,48	0.17%	14,57,51	0.00%	0.01%
10,0.5	1.5	25	35.19	97.7%	0.01%	28,48	0.39%	27,28	0.83%	28,49,29	0.23%	0.31%
10,0.5	1.5	50	43.28	97.3%	0.01%	27,57	0.08%	26,36	0.31%	26,57,40	0.01%	0.02%
10,0.5	1.5	100	54.89	97.5%	0.00%	25,68	0.05%	25,48	0.16%	25,68,53	0.00%	0.01%
10,0.5	2.5	25	36.77	97.3%	2.60%	39,59	0.95%	39,25	1.17%	39,59,28	0.88%	0.91%
10,0.5	2.5	50	44.89	97.0%	0.04%	37,68	0.40%	36,37	0.40%	37,69,38	0.23%	0.33%
10,0.5	2.5	100	56.43	96.8%	0.03%	35,79	0.05%	35,49	0.14%	35,79,54	0.01%	0.02%
10,0.5	3.5	25	37.98	97.1%	8.37%	49,69	1.27%	49,25	1.24%	49,70,29	1.15%	1.19%
10,0.5	3.5	50	46.01	96.8%	1.91%	48,77	1.06%	48,35	1.18%	48,78,37	1.02%	1.05%
10,0.5	3.5	100	57.72	96.7%	0.00%	46,91	0.14%	46,47	0.09%	46,91,55	0.10%	0.11%

Table 6.15: The results for the exact cost model when the demand follows a negative binomial distribution.

w, u	L	K	s', Q'	$C_{s',Q'}$	$\tilde{s}_1, \tilde{S}_1$	$C_{\tilde{s}_1,\tilde{S}_1}$	$\tilde{s}_2, \tilde{S}_2$	$C_{\tilde{s}_2,\tilde{S}_2}$	s_{BO}, S_{BO}	$C_{s_{BO},S_{BO}}$	$C_{\tilde{s}_1,\tilde{S}_1,\tilde{q}_1}$	$C_{\tilde{s}_2,\tilde{S}_2,\tilde{q}_2}$	$C_{s_{BO},S_{BO},q_{BO}}$
2,0.5	0.5	25	6,11	9.46%	3,13	0.02%	3,13	0.02%	3,14	0.13%	0.05%	0.05%	0.03%
2,0.5	0.5	50	6,15	10.19%	2,17	0.03%	2,17	0.03%	2,17	0.03%	0.00%	0.00%	0.00%
2,0.5	0.5	100	0,22	1.75%	1,22	0.17%	1,22	0.17%	2,22	0.01%	0.17%	0.17%	0.00%
2,0.5	1.5	25	10,11	16.24%	5,16	0.16%	5,16	0.16%	5,17	0.56%	0.00%	0.00%	0.22%
2,0.5	1.5	50	0,20	11.80%	4,19	0.11%	4,19	0.11%	5,20	0.26%	0.03%	0.03%	0.08%
2,0.5	1.5	100	0,23	4.47%	2,23	0.87%	2,24	0.97%	3,25	0.32%	0.91%	0.83%	0.04%
2,0.5	2.5	25	14,11	24.27%	6,17	1.81%	7,19	0.43%	8,19	0.25%	2.09%	0.09%	0.04%
2,0.5	2.5	50	1,20	12.66%	5,20	1.25%	6,22	0.34%	7,23	0.58%	1.64%	0.02%	0.27%
2,0.5	2.5	100	0,24	6.88%	3,23	1.85%	4,26	0.75%	5,27	0.38%	2.18%	0.51%	0.03%
2,0.5	3.5	25	17,11	26.76%	8,19	1.90%	9,21	0.45%	10,22	0.59%	2.26%	0.17%	0.31%
2,0.5	3.5	50	1,21	16.69%	6,21	2.94%	8,24	0.30%	9,25	0.51%	3.60%	0.01%	0.21%
2,0.5	3.5	100	0,25	8.90%	4,18	8.77%	6,29	0.86%	7,30	0.73%	11.59%	0.38%	0.16%
2,2/7	0.5	25	18,18	19.49%	9,25	0.08%	9,25	0.08%	10,26	0.09%	0.07%	0.07%	0.01%
2,2/7	0.5	50	18,24	19.28%	8,30	0.04%	8,30	0.04%	8,31	0.06%	0.06%	0.06%	0.01%
2,2/7	0.5	100	17,33	15.96%	6,38	0.14%	6,38	0.14%	7,39	0.05%	0.13%	0.13%	0.00%
2,2/7	1.5	25	28,18	28.33%	14,31	0.75%	15,32	0.19%	16,33	0.25%	0.69%	0.04%	0.06%
2,2/7	1.5	50	27,24	25.11%	12,36	0.94%	13,37	0.35%	14,38	0.23%	0.90%	0.19%	0.04%
2,2/7	1.5	100	26,33	21.43%	10,43	0.83%	11,44	0.32%	12,46	0.28%	0.80%	0.21%	0.06%
2,2/7	2.5	25	37,18	35.62%	18,36	2.52%	20,39	0.82%	21,40	0.81%	2.58%	0.58%	0.55%
2,2/7	2.5	50	35,25	29.02%	16,41	2.14%	18,43	0.51%	20,45	0.42%	2.12%	0.31%	0.19%
2,2/7	2.5	100	0,50	24.30%	13,47	2.50%	16,51	0.48%	17,52	0.36%	2.77%	0.23%	0.08%
2,2/7	3.5	25	45,18	41.26%	23,41	2.88%	26,45	1.56%	27,46	1.86%	2.98%	1.77%	2.04%
2,2/7	3.5	50	43,25	33.38%	20,45	3.61%	24,50	0.43%	25,51	0.48%	3.92%	0.20%	0.23%
2,2/7	3.5	100	1,52	27.32%	16,50	4.56%	21,57	0.53%	23,58	0.38%	5.44%	0.25%	0.14%
10,0.5	0.5	25	25,25	16.22%	15,35	0.94%	15,35	0.94%	17,37	0.07%	0.97%	0.97%	0.01%
10,0.5	0.5	50	25,33	15.22%	14,43	0.48%	14,43	0.48%	15,45	0.06%	0.48%	0.48%	0.00%
10,0.5	0.5	100	24,46	12.21%	12,55	0.57%	12,55	0.57%	14,57	0.05%	0.58%	0.58%	0.00%
10,0.5	1.5	25	40,25	22.75%	25,46	2.82%	27,48	0.60%	28,50	0.52%	2.87%	0.55%	0.45%
10,0.5	1.5	50	40,33	20.87%	24,54	1.38%	25,56	0.44%	27,57	0.08%	1.43%	0.39%	0.02%
10,0.5	1.5	100	39,46	17.28%	21,65	2.01%	23,67	0.42%	25,69	0.08%	2.08%	0.37%	0.01%
10,0.5	2.5	25	55,25	30.92%	35,56	3.77%	38,60	1.22%	39,62	1.44%	3.83%	1.22%	1.42%
10,0.5	2.5	50	54,34	25.52%	33,64	3.49%	36,68	0.61%	38,69	0.53%	3.60%	0.56%	0.48%
10,0.5	2.5	100	52,42	19.62%	30,74	3.66%	34,79	0.25%	36,81	0.14%	3.94%	0.20%	0.07%
10,0.5	3.5	25	68,25	34.64%	44,66	5.60%	49,72	1.67%	51,73	2.06%	5.59%	1.50%	1.89%
10,0.5	3.5	50	67,27	29.73%	41,73	6.37%	47,80	1.30%	49,81	1.48%	6.36%	1.36%	1.47%
10,0.5	3.5	100	64,30	26.71%	38,82	7.15%	45,91	0.20%	46,92	0.18%	7.99%	0.15%	0.11%

Table 6.16: The results for the approximate cost model when the demand follows a negative binomial distribution.

λ, μ	L	$\overline{\beta}$	C^*	$\hat{s}, \hat{Q}$	$C_{\hat{s},\hat{Q}}$	$\bar{s}, \bar{S}$	$C_{\bar{s},\bar{S}}$	s^*, S^*, q^*	$C_{s*,S*,q*}$	$C_{\bar{s},\bar{S},\bar{q}}$	$\beta_{\bar{s},\bar{S},\bar{q}}$
2,1	0.5	85%	12.45	0,12	2.15%	0,12	2.15%	0,12,12	2.15%	2.15%	85.7%
2,1	0.5	90%	13.25	0,18	2.24%	0,18	2.24%	0,18,18	2.24%	2.24%	90.0%
2,1	0.5	95%	14.34	2,14	1.94%	2,15	1.89%	2,15,14	1.84%	1.84%	95.9%
2,1	1.5	85%	12.52	2,13	1.74%	2,14	1.81%	2,14,13	1.62%	1.53%	85.0%
2,1	1.5	90%	13.44	3,13	1.36%	3,15	1.48%	3,15,13	1.34%	1.50%	89.7%
2,1	1.5	95%	14.74	4,16	0.45%	4,19	0.47%	4,19,16	0.00%	0.00%	95.0%
2,1	2.5	85%	12.61	4,13	1.00%	4,16	1.21%	4,17,13	1.00%	0.71%	84.2%
2,1	2.5	90%	13.64	5,14	0.29%	5,18	0.43%	5,19,14	0.29%	-0.14%	89.4%
2,1	2.5	95%	15.15	6,19	1.96%	6,24	1.99%	6,24,20	1.67%	1.22%	94.9%
5,1	0.5	85%	19.46	1,23	0.14%	1,23	0.00%	1,23,23	0.00%	-0.56%	84.6%
5,1	0.5	90%	20.76	3,21	0.72%	3,22	0.76%	3,22,21	0.66%	0.64%	89.7%
5,1	0.5	95%	22.58	5,21	1.06%	5,24	1.04%	5,25,22	0.97%	0.98%	95.2%
5,1	1.5	85%	19.58	6,24	0.43%	6,28	1.08%	6,29,24	0.25%	-1.13%	84.1%
5,1	1.5	90%	21.11	8,23	0.12%	8,29	0.39%	8,31,23	0.12%	-0.69%	89.5%
5,1	1.5	95%	23.21	10,27	0.99%	10,34	0.64%	10,34,29	0.42%	0.04%	94.9%
5,1	2.5	85%	19.75	11,25	0.66%	11,33	0.47%	11,34,25	0.10%	-1.91%	83.6%
5,1	2.5	90%	21.40	13,26	0.86%	13,36	0.61%	13,37,26	0.32%	-0.93%	89.5%
5,1	2.5	95%	23.78	16,23	0.23%	16,37	0.36%	16,37,25	0.15%	0.06%	95.0%
10,1	0.5	85%	27.40	5,29	0.53%	4,36	0.69%	4,37,35	0.40%	-1.23%	84.1%
10,1	0.5	90%	29.37	7,32	0.27%	7,35	0.42%	7,35,33	0.11%	-0.39%	89.5%
10,1	0.5	95%	32.19	10,33	0.08%	10,38	0.00%	10,40,33	0.16%	-0.39%	94.9%
10,1	1.5	85%	27.63	15,30	0.41%	15,41	0.95%	15,43,31	0.36%	0.13%	84.3%
10,1	1.5	90%	29.85	17,36	0.35%	17,48	0.91%	17,52,36	0.27%	-0.84%	89.5%
10,1	1.5	95%	32.96	21,32	0.45%	21,48	0.35%	21,50,32	0.20%	0.11%	95.0%
10,1	2.5	85%	27.90	24,36	0.30%	25,51	0.65%	25,56,35	0.10%	2.49%	85.8%
10,1	2.5	90%	30.32	28,36	0.27%	28,56	1.35%	28,58,36	0.13%	2.21%	91.0%
10,1	2.5	95%	33.60	31,32	0.16%	31,65	1.27%	31,61,32	0.12%	-0.01%	94.7%
2.5,2	0.5	85%	20.49	2,24	0.17%	2,25	0.61%	2,25,24	0.00%	-0.67%	84.6%
2.5,2	0.5	90%	22.23	4,25	0.29%	4,27	0.60%	4,28,25	0.12%	-0.17%	89.9%
2.5,2	0.5	95%	24.97	7,26	0.74%	7,30	0.64%	7,30,27	0.13%	0.13%	95.0%
2.5,2	1.5	85%	21.03	8,23	0.70%	8,28	1.14%	8,29,23	0.28%	-0.66%	84.0%
2.5,2	1.5	90%	23.18	10,26	0.92%	10,32	0.52%	10,32,29	0.15%	-0.77%	89.6%
2.5,2	1.5	95%	26.57	14,24	0.76%	14,35	0.80%	14,35,26	0.33%	0.42%	95.1%
2.5,2	2.5	85%	21.60	13,25	0.01%	13,35	0.85%	13,36,26	0.05%	-1.45%	84.2%
2.5,2	2.5	90%	24.05	16,25	0.18%	16,38	0.71%	16,39,26	0.05%	-0.42%	89.8%
2.5,2	2.5	95%	27.97	20,26	0.16%	20,43	0.31%	20,43,29	0.18%	-0.18%	95.0%

Table 6.17: The results for the exact service model when the demand follows a pure or compound Poisson process.

λ,μ	L	$\overline{\beta}$	s',S'	$C_{s',S'}$	$\tilde{s_1},\tilde{S_1}$	$C_{\tilde{s_1},\tilde{S_1}}$	$\tilde{s_2},\tilde{S_2}$	$C_{\tilde{s_2},\tilde{S_2}}$	s_{BO},S_{BO}	$C_{s_{BO},S_{BO}}$	$C_{\tilde{s_1},\tilde{S_1},\tilde{q_1}}$	$C_{\tilde{s_2},\tilde{S_2},\tilde{q_2}}$	$C_{s_{BO},S_{BO},q_{BO}}$
2,1	0.5	85%	0,14	2.94%	0,12	2.15%	0,12	2.15%	0,13	2.31%	2.15%	2.15%	2.31%
2,1	0.5	90%	0,14	-3.32%	0,18	2.24%	0,18	2.24%	1,14	2.99%	2.24%	2.24%	2.89%
2,1	0.5	95%	2,16	2.26%	2,15	1.89%	2,15	1.89%	2,15	1.89%	1.84%	1.84%	1.84%
2,1	1.5	85%	1,15	-2.88%	1,19	4.42%	1,20	6.72%	2,16	3.43%	3.14%	5.36%	2.43%
2,1	1.5	90%	3,17	2.55%	2,20	2.70%	2,21	4.79%	3,16	1.81%	0.99%	2.94%	1.36%
2,1	1.5	95%	4,18	-0.38%	4,17	-0.89%	4,17	-0.89%	4,20	1.61%	-1.07%	-1.07%	1.03%
2,1	2.5	85%	3,17	-2.86%	3,21	4.74%	3,22	7.09%	4,18	3.02%	2.49%	4.70%	1.73%
2,1	2.5	90%	5,19	1.25%	4,22	1.82%	5,18	0.43%	5,20	2.39%	-0.25%	-0.14%	1.12%
2,1	2.5	95%	6,20	-3.02%	6,19	-3.59%	6,23	0.42%	6,25	3.73%	-3.82%	-0.28%	2.90%
5,1	0.5	85%	1,23	0.00%	1,22	-0.67%	1,22	-0.67%	1,26	2.77%	-1.11%	-1.11%	1.92%
5,1	0.5	90%	2,24	-2.22%	2,26	-0.64%	2,26	-0.64%	3,24	1.40%	-1.84%	-1.84%	0.91%
5,1	0.5	95%	5,27	1.85%	4,28	-0.91%	4,28	-0.91%	5,25	1.14%	-1.51%	-1.51%	1.00%
5,1	1.5	85%	6,28	1.08%	5,30	0.91%	6,27	0.20%	6,32	5.64%	-2.06%	-1.69%	2.45%
5,1	1.5	90%	8,30	1.06%	7,28	-3.43%	7,33	1.33%	8,32	2.76%	-5.00%	-1.56%	1.07%
5,1	1.5	95%	10,32	-0.69%	9,33	-3.05%	10,30	-1.54%	10,36	2.38%	-4.05%	-1.83%	1.66%
5,1	2.5	85%	11,33	0.47%	10,31	-3.96%	11,34	1.49%	11,38	6.44%	-6.26%	-1.18%	2.91%
5,1	2.5	90%	13,35	-0.20%	11,38	-1.78%	13,35	-0.20%	13,39	3.64%	-5.04%	-1.52%	1.63%
5,1	2.5	95%	16,38	0.84%	14,38	-5.21%	15,41	0.11%	16,38	0.84%	-6.22%	-0.60%	0.46%
10,1	0.5	85%	3,35	-2.21%	3,37	-0.65%	3,37	-0.65%	4,40	3.96%	-2.55%	-2.55%	1.35%
10,1	0.5	90%	6,38	-0.51%	6,33	-2.85%	6,33	-2.85%	7,38	1.77%	-3.49%	-3.49%	0.48%
10,1	0.5	95%	10,42	1.53%	9,36	-3.11%	9,36	-3.11%	10,40	0.66%	-3.56%	-3.56%	0.13%
10,1	1.5	85%	14,46	1.85%	12,43	-3.64%	13,47	1.26%	14,51	6.44%	-7.83%	-3.40%	2.30%
10,1	1.5	90%	17,49	1.52%	15,42	-6.18%	16,47	-1.67%	17,51	2.94%	-7.15%	-3.89%	0.62%
10,1	1.5	95%	20,52	-0.56%	18,52	-4.44%	20,48	-2.10%	21,50	0.96%	-6.08%	-2.39%	0.72%
10,1	2.5	85%	24,56	0.99%	21,50	-8.22%	24,56	0.99%	25,58	4.28%	-13.61%	-2.66%	0.61%
10,1	2.5	90%	27,59	-0.10%	24,51	-9.99%	27,57	-1.18%	28,59	1.99%	-10.28%	-2.75%	-0.11%
10,1	2.5	95%	31,63	0.26%	28,57	-8.54%	31,59	-0.20%	32,60	2.61%	-8.57%	0.46%	3.10%
2.5,2	0.5	85%	2,24	-0.28%	2,24	-0.28%	2,24	-0.28%	2,27	2.70%	-1.38%	-1.38%	1.11%
2.5,2	0.5	90%	4,26	-0.19%	4,25	-0.86%	4,25	-0.86%	4,28	1.49%	-1.42%	-1.42%	0.63%
2.5,2	0.5	95%	7,29	-0.01%	7,26	-1.26%	7,26	-1.26%	7,30	0.64%	-1.41%	-1.41%	0.13%
2.5,2	1.5	85%	7,29	-0.84%	7,28	-1.81%	7,31	1.37%	8,31	3.92%	-3.94%	-2.16%	1.34%
2.5,2	1.5	90%	10,32	0.52%	9,32	-2.23%	9,35	0.93%	10,35	3.36%	-4.06%	-1.30%	1.74%
2.5,2	1.5	95%	14,36	1.36%	12,38	-2.40%	13,36	-1.51%	14,36	1.36%	-3.51%	-2.17%	0.92%
2.5,2	2.5	85%	13,35	0.85%	11,37	-0.98%	12,39	3.57%	13,39	5.55%	-5.69%	-0.28%	2.67%
2.5,2	2.5	90%	16,38	0.71%	14,37	-5.28%	15,41	1.18%	16,41	3.52%	-7.40%	-0.73%	2.15%
2.5,2	2.5	95%	20,42	-0.30%	18,41	-6.34%	19,45	-0.65%	20,44	1.01%	-7.07%	-1.29%	0.48%

Table 6.18: The results for the approximate service model when the demand follows a pure or compound Poisson process.

λ, μ	L	$\overline{\beta}$	C^*	$\hat{s}, \hat{Q}$	$C_{\hat{s},\hat{Q}}$	$\bar{s}, \bar{S}$	$C_{\bar{s},\bar{S}}$	s^*, S^*, q^*	$C_{s*,S*,q*}$	$C_{\bar{s},\bar{S},\bar{q}}$	$\beta_{\bar{s},\bar{S},\bar{q}}$
2,0.5	0.5	85%	12.87	0,15	0.97%	0,15	0.97%	0,15,15	0.97%	0.97%	85.7%
2,0.5	0.5	90%	13.83	1,16	0.97%	1,16	0.49%	1,16,16	0.49%	-0.21%	89.9%
2,0.5	0.5	95%	15.37	3,14	1.27%	3,15	1.04%	3,15,15	1.04%	1.05%	95.0%
2,0.5	1.5	85%	13.07	2,16	1.11%	2,17	1.19%	2,17,16	0.53%	-0.39%	84.8%
2,0.5	1.5	90%	14.28	3,18	1.14%	3,20	1.62%	3,21,18	1.14%	-0.53%	89.6%
2,0.5	1.5	95%	16.15	5,19	1.52%	6,18	2.06%	5,23,20	1.42%	1.98%	95.2%
2,0.5	2.5	85%	13.32	4,17	1.24%	4,20	2.11%	4,20,17	0.56%	-0.51%	84.7%
2,0.5	2.5	90%	14.67	6,15	0.20%	6,20	0.81%	6,20,16	0.13%	-0.22%	89.8%
2,0.5	2.5	95%	16.80	8,17	0.53%	8,24	1.05%	8,24,17	0.02%	0.64%	95.2%
2,2/7	0.5	85%	20.75	3,21	0.81%	2,26	0.52%	2,26,26	0.52%	-0.95%	84.4%
2,2/7	0.5	90%	22.65	5,22	0.57%	4,29	0.79%	5,25,23	0.46%	-0.18%	89.8%
2,2/7	0.5	95%	25.69	8,24	0.51%	8,29	0.39%	8,29,27	0.19%	0.08%	95.0%
2,2/7	1.5	85%	21.40	8,25	0.87%	8,29	0.47%	8,29,28	0.26%	-1.77%	83.7%
2,2/7	1.5	90%	23.75	10,28	0.56%	11,31	0.55%	11,32,26	0.39%	-0.35%	89.6%
2,2/7	1.5	95%	27.50	14,29	0.70%	15,36	0.90%	14,40,32	0.40%	0.51%	95.1%
2,2/7	2.5	85%	22.05	13,27	0.32%	14,34	0.93%	13,38,28	0.25%	-0.81%	84.2%
2,2/7	2.5	90%	24.73	17,25	1.09%	17,38	1.00%	16,41,32	0.28%	0.03%	89.8%
2,2/7	2.5	95%	29.04	21,27	0.39%	21,44	0.01%	21,45,29	0.10%	-0.49%	94.9%
10,0.5	0.5	85%	28.00	5,33	0.23%	5,35	0.62%	5,36,33	0.05%	-0.83%	84.3%
10,0.5	0.5	90%	30.24	8,32	0.35%	8,35	0.31%	8,38,32	0.21%	-0.33%	89.3%
10,0.5	0.5	95%	33.55	11,37	0.49%	11,43	0.50%	11,46,37	0.27%	-0.19%	94.9%
10,0.5	1.5	85%	28.50	16,31	0.44%	15,45	0.76%	15,48,36	0.22%	-2.01%	83.6%
10,0.5	1.5	90%	31.14	19,32	0.38%	19,46	0.75%	19,47,34	0.25%	-0.24%	89.6%
10,0.5	1.5	95%	35.07	23,35	0.61%	23,52	0.32%	23,53,37	0.13%	-0.23%	94.9%
10,0.5	2.5	85%	29.06	26,35	0.34%	26,54	0.64%	26,58,35	0.06%	-0.69%	84.3%
10,0.5	2.5	90%	31.99	29,38	0.60%	29,60	0.40%	29,61,41	0.21%	-1.09%	89.5%
10,0.5	2.5	95%	36.24	34,35	1.34%	34,62	1.04%	34,63,40	0.91%	1.16%	95.1%

Table 6.19: The results for the exact service model when the demand follows a negative binomial distribution.

λ,μ	L	$\overline{\beta}$	s',S'	$C_{s',S'}$	$\tilde{s}_1,\tilde{S}_1$	$C_{\tilde{s}_1,\tilde{S}_1}$	$\tilde{s}_2,\tilde{S}_2$	$C_{\tilde{s}_2,\tilde{S}_2}$	s_{BO},S_{BO}	$C_{s_{BO},S_{BO}}$	$C_{\tilde{s}_1,\tilde{S}_1,\tilde{q}_1}$	$C_{\tilde{s}_2,\tilde{S}_2,\tilde{q}_2}$	$C_{s_{BO},S_{BO},q_{BO}}$
2,0.5	0.5	85%	0,14	-0.20%	0,15	0.97%	0,15	0.97%	0,16	2.44%	0.97%	0.97%	2.44%
2,0.5	0.5	90%	1,15	-0.58%	1,16	0.49%	1,16	0.49%	1,17	1.84%	-0.21%	-0.21%	0.97%
2,0.5	0.5	95%	3,17	1.76%	3,15	1.04%	3,15	1.04%	3,16	1.21%	1.05%	1.05%	1.11%
2,0.5	1.5	85%	2,16	-0.30%	2,17	1.19%	2,18	2.93%	2,19	4.87%	-0.39%	1.11%	2.85%
2,0.5	1.5	90%	3,17	-2.88%	3,19	-0.09%	3,20	1.62%	4,17	2.40%	-2.00%	-0.53%	1.84%
2,0.5	1.5	95%	5,19	-2.96%	5,20	-2.00%	5,22	0.60%	6,19	2.32%	-2.52%	-0.08%	2.15%
2,0.5	2.5	85%	4,18	-1.30%	4,20	2.11%	4,21	4.10%	4,22	6.25%	-0.51%	1.24%	3.19%
2,0.5	2.5	90%	6,20	0.81%	5,23	1.56%	6,20	0.81%	6,22	3.54%	-1.25%	-0.22%	2.13%
2,0.5	2.5	95%	8,22	-1.04%	7,26	0.25%	8,23	-0.11%	8,24	1.05%	-0.67%	-0.47%	0.64%
2,2/7	0.5	85%	2,24	-1.40%	2,26	0.52%	2,26	0.52%	2,28	2.80%	-0.95%	-0.95%	1.07%
2,2/7	0.5	90%	4,26	-1.91%	4,27	-1.11%	4,27	-1.11%	5,26	1.40%	-1.90%	-1.90%	0.93%
2,2/7	0.5	95%	8,30	0.94%	7,31	-1.34%	7,31	-1.34%	8,30	0.94%	-1.96%	-1.96%	0.57%
2,2/7	1.5	85%	7,29	-2.24%	7,31	-0.01%	7,33	2.52%	8,33	4.81%	-3.65%	-1.59%	1.68%
2,2/7	1.5	90%	10,32	-1.67%	9,35	-1.11%	10,33	-0.78%	11,34	2.84%	-3.48%	-2.34%	1.61%
2,2/7	1.5	95%	15,37	1.46%	13,38	-3.05%	14,36	-1.98%	14,41	2.11%	-3.98%	-2.55%	1.31%
2,2/7	2.5	85%	13,35	-0.77%	12,35	-3.03%	12,41	4.93%	13,41	6.72%	-6.46%	0.56%	3.34%
2,2/7	2.5	90%	16,38	-1.75%	14,41	-3.09%	16,40	0.09%	16,44	4.53%	-5.96%	-1.47%	2.70%
2,2/7	2.5	95%	21,43	-0.58%	19,43	-5.70%	20,46	-0.88%	21,46	1.48%	-6.46%	-1.52%	0.90%
10,0.5	0.5	85%	4,36	-0.95%	4,35	-1.64%	4,35	-1.64%	5,38	2.67%	-3.45%	-3.45%	0.76%
10,0.5	0.5	90%	7,39	-0.13%	6,40	-1.50%	6,40	-1.50%	8,38	1.56%	-3.17%	-3.17%	0.44%
10,0.5	0.5	95%	11,43	0.50%	10,41	-2.70%	10,41	-2.70%	11,45	1.64%	-3.48%	-3.48%	0.84%
10,0.5	1.5	85%	14,46	-0.18%	13,43	-4.13%	14,46	-0.18%	15,51	5.79%	-7.96%	-3.94%	2.00%
10,0.5	1.5	90%	18,50	0.83%	16,45	-6.00%	17,50	-0.95%	18,54	3.84%	-7.94%	-3.41%	1.55%
10,0.5	1.5	95%	22,54	-0.87%	20,53	-5.23%	21,56	-1.56%	23,54	1.19%	-6.58%	-2.74%	0.53%
10,0.5	2.5	85%	25,57	0.56%	22,51	-8.61%	24,61	2.91%	25,64	7.20%	-12.59%	-1.12%	3.04%
10,0.5	2.5	90%	29,61	0.91%	25,56	-9.37%	28,62	-0.27%	29,65	3.76%	-12.54%	-2.59%	1.87%
10,0.5	2.5	95%	34,66	1.61%	30,63	-8.09%	33,66	-0.66%	34,68	2.36%	-9.60%	-1.45%	1.68%

Table 6.20: The results for the approximate service model when the demand follows a negative binomial distribution.

Chapter 7

Lost sales in hospitals

From the previous two chapters we concluded that restricted order-up-to policies are near-optimal replenishment policies which are simple to implement and use in practice, whereas optimal policies are difficult to compute. We developed general procedures to approximate the steady-state behavior of lost-sales inventory systems and determined near-optimal values of the inventory control variables. However, when specific characteristics or extra requirements of an inventory system are known, they should be included in such (approximation) procedures. In this chapter we illustrate this principle for inventory systems at hospitals.

The inventory management and control at hospitals should be simple and effective to ensure the availability of items. It is not the main objective for a hospital to minimize the total inventory costs. Their main concern is the availability of items, otherwise the demand is lost since medical care is urgent, and the hospital staff cannot wait for backorders. They either find the same item at a different storage location, or they take a different item which is available. Another characteristic of a hospital inventory control system is that the replenishment process should be simple. Hospital staff does not want to spend time on inventory control. They only want to indicate whether more units of an item are required. The actual order size is not relevant for them. To speed up the replenishment process, order sizes are fixed such that the remaining units on stock are not counted at an order instant. Consequently, fixed order size policies (FOSPs) are required for such situations. This is a clear example in which the requirements of a practical setting prohibit to apply (restricted) order-up-to policies in a lost-sales setting, even though such policies perform better (see Chapter 5 and Chapter 6). Furthermore, hospital departments have to deal with storage capacity limitations at each care unit. The goal of this chapter is to develop a general inventory control model for hospitals which considers the service level as well as the capacity restrictions. The model is a simplification in comparison to the models developed in previous chapters, due to the specific requirements for a hospital environment. Based on this mathematical model the reorder levels and order quantities are determined for the most common replenishment policy at hospitals (i.e., FOSP). Since the available storage capacity at hospitals is limited, we also embed our single-item model in a multi-item inventory control system with capacity constraints. Our second goal is to derive a simple and synoptic inventory rule that can be used by hospital staff to set the values of these inventory control variables. Since the possible values are restricted due to the limitations in storage capacity, our approximation procedure should incorporate this constraint.

Section 7.1 gives a more detailed introduction to the daily practices regarding inventory management in hospitals. In Section 7.2, the specific characteristics of the inventory replenishment system under study are provided, and these characteristics are related to the existing literature on inventory control in hospitals. A new inventory model is developed in Section 7.3 that satisfies the common characteristics as described in Section 7.2. Moreover, we present an approach to use the single-item model in a multi-item system to assign items to the available capacity at hospitals. A disadvantage of the most commonly used replenishment policy is that the available storage capacity is not used efficiently since fixed order sizes are required. The available capacity is utilized more efficiently when the replenishment policy prescribes to raise the inventory position to the available capacity at each review. Consequently, higher service levels are obtained with such policies. Therefore, we consider two order-up-to policies as alternative replenishment policies throughout this chapter which have already been studied in Chapter 5 and Chapter 6. Although such order-up-to policies require more effort to process the replenishments, we show the benefits of such policies in terms of service level and capacity utilization compared to the fixed order size policy which is commonly used in practice. In Section 7.4 we demonstrate and compare the performance of the different replenishment policies based on a case study. It requires an exhaustive search procedure to find optimal values of the inventory control variables for the policies. Therefore, a new heuristic approach is formulated in Section 7.5 to derive near-optimal solutions indicating the reorder level and order quantity. The requirements for such an inventory rule are that it should be insightful and synoptic for hospitals, such that it can be easily applied by hospital staff. Numerical experiments are performed in Section 7.6 to test the performance of the different replenishment policies, and the inventory rule for a wide range of settings. The conclusions are presented in Section 7.7.

7.1 Inventory management in hospitals

The main objective of a hospital is to provide high-quality health care. Sufficient medical items need to be on hand to enable hospital staff to perform their daily work. Typically, medical supplies are stored at many locations in a hospital and in large quantities to ensure that stock outs hardly occur. As a result, hospitals have lack of storage space (Lapierre and Ruiz [173], Little and Coughlan [184]) and millions of dollars are tied up in inventories that consume on average 20% of net patients revenues and represent the second largest expense after labor (Moon [208]). Therefore, it is important to find a balance between the desired service quality and the required inventory levels.

Traditionally, health care supply chains are characterized by a multitude of different suppliers, products and patient care units that arbitrarily order multiple items (Rivard-Rover et al. [247]). A hospital storage room receives incoming units from the suppliers and distributes the units to lower-level point-of-use (POU) locations, such as nursing units and operating rooms, based on orders placed by the POU locations. This process is depicted in Figure 7.1a. Another possibility is to outsource such replenishment activities or for suppliers to deliver the items directly to the POU locations (see Figure 7.1b). Such stockless or just-in-time (JIT) inventory systems are described by, for example, Nathan and Trinkaus [220] and Danas et al. [66]. The cost and service level differences between the traditional and the JIT stockless inventory systems are quantified by Nicholson et al. [224].

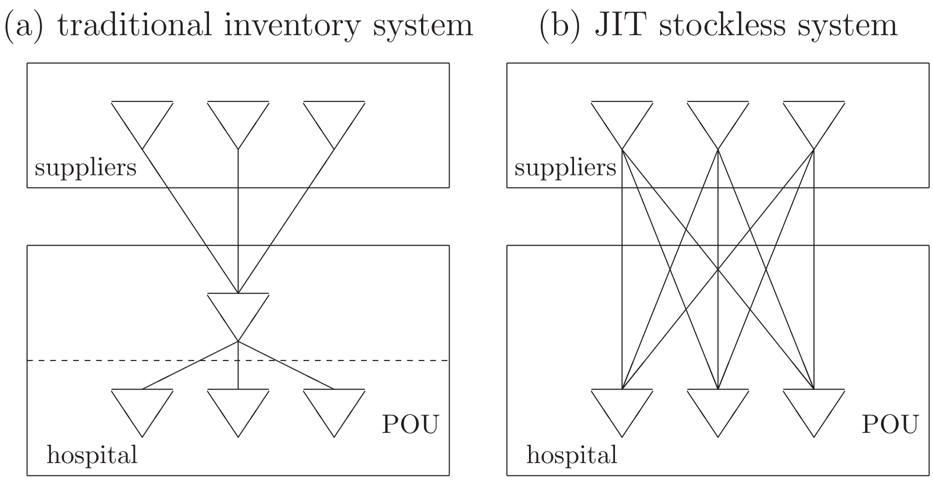


Figure 7.1: A traditional inventory system versus a JIT stockless system.

In the literature two concepts are proposed to improve the performance of health care supply chains: product standardization and selecting prime vendors (see, e.g., Johnston [140]). The former concept reduces the number of different items that have to be stocked. Therefore, it reduces the volatility of the demand. The latter concept can reduce lead times and, therefore, the safety stock as well. Such decisions regarding standardization and the choice of suppliers, are on tactical and strategical level. However, it remains essential to decide upon the order moment and order size for each item on stock at a single POU location (decisions at operational level). This is the topic for the remainder of this chapter. In the next section we describe the specific characteristics of the replenishment process in current practices at hospitals, and relate them to the literature on inventory control in hospitals.

7.2 Replenishment process: practice vs literature

We can distinguish roughly three types of inventories in hospitals, namely perishable items including medicines and blood (see, e.g., Katsaliaki and Brailsford 2007), non-disposables (e.g., instruments) and disposables (e.g., gloves, needles, sutures). Our main focus in this research is on disposable items since these products are stored all over the hospital and, therefore, more difficult to control.

The logistics processes related to the inventory management of disposable items at each of the POU locations can be described as an inventory system where all items are stored in bins. Each bin has a total storage capacity of C units for a particular item that can be used to fulfill demand of medical staff whenever required. If s or less units of an item are available in the bin a signal is given (e.g., by putting a bar code of the item on an ordering board). An employee of the replenishment department scans these signals to complete the actual order at prespecified time intervals of length R (i.e., periodic reviews) that may range from days to weeks. Items are usually ordered in fixed quantities Q to provide a transparent and insightful inventory policy for hospital staff. This replenishment process corresponds to the fixed order size policy (FOSP) as described in Section 6.2.4. After L time units the ordered items are resupplied from, for example, a central storage room to the specific bins. In a hospital setting the lead time L is known and relatively short due to the high product availability at nearby higher-level stock points. Therefore, another characteristic for the inventory system is that the lead time L is shorter than the length of the review period R (see also Duclos [76]). This is referred to as *fractional lead times.*

As mentioned before, a patient always needs to receive the required service in the health care sector. If a required item is not available in the right quantity at a specific POU location, a substitute product is used or an emergency delivery is performed (e.g., from another POU location). The original demand for the item is considered to be lost. Such situations are time costly and should be avoided. Therefore, we define the service level as the fraction of demand to be satisfied directly from stock on hand (i.e., *fill rate*). To summarize, the inventory system in a hospital is characterized by periodic reviews, an (R, s, Q) replenishment policy (or FOSP), short lead times, lost sales, capacity restrictions and a service level objective.

Hardly any literature is available on replenishment policies for inventory systems in a hospital setting. The literature overview on operation research (OR) in health care settings by Brandeau et al. [45] does not even mention inventory theory. Only a few papers are available about inventory control in a hospital setting. Lapierre and Ruiz [173] solve a multi-item inventory replenishment problem with storage and manpower capacity restrictions. In their research, demand is assumed to be deterministic and known. They formulate a non-linear mixed-integer problem and solve this with a tabu search metaheuristic. Similar restrictions are considered by Little and Coughlan [184]. The authors propose an optimization model based on constraint programming to determine the delivery frequency and order sizes. A single item inventory problem with capacity restrictions is discussed by Vincent and Ranton [299] for a hospital environment. They extend the basic EOQ formula and focus on a cost objective instead of a service objective. Order costs are also the main focus in Dellaert and Van de Poel [72], where joint replenishments at the central storage room in a hospital result in cost savings. All papers mentioned so far do not correspond to a model with similar characteristics for the replenishment policy as commonly observed at POU locations in hospitals. To our knowledge, we are the first authors to study a lost-sales inventory system with capacity restrictions and a fixed order size policy. Downs et al. [74] perform a similar study but for base-stock policies. Therefore, we develop a new inventory control model in the next section.

7.3 Model

According to the characteristics described in Section 7.2, we model the most common situation seen at hospitals as a periodic review inventory model with lost sales and an (R, s, Q) replenishment policy. The lead time is fractional compared to the review period and the maximum inventory position should not exceed the available storage capacity (i.e., $s+Q \leq C$). Such replenishment policies are simple to implement, but they do not use the available capacity efficiently. Order-up-to (R, s, S) policies incorporate flexible order sizes and, therefore, use the capacity more efficiently. To illustrate the impact on the service level for such policies we also model this replenishment policy. Notice that the (R, s, Q) and (R, s, S) replenishment policies correspond to the FOSP and POUTP of Chapter 6, respectively. A model for the (R, s, S) type of policies with similar characteristics (lost sales, fractional lead times, and a service level constraint) is developed by Kapalka et al. [154]. They propose a procedure to determine the service level based on the steady-state distribution of the on-hand inventory at a review. We use a similar approach in this section for the (R, s, Q) policy. Notice that such single-item models are simplifications compared to the models developed in Section 6.2, since fractional lead times are assumed in this chapter. After we present these single-item models, we embed the models in a multi-item inventory system with capacity restrictions.

7.3.1 Single-item inventory system

In Chapter 5 and Chapter 6 the behavior of the inventory system is described by a multi-dimensional Markov chain at review time T and order delivery time t. However, in this chapter an order is delivered within the same review period as it is ordered in. Therefore, the inventory system can be modeled as a one-dimensional Markov chain, where we only consider the inventory status at a review instant. The demand during a period of length τ is modeled as a discrete random variable D_τ, which is assumed to be independent for non-overlapping time intervals. The probability distribution function is given by $g_\tau(d)$ and $\mathcal{G}^0_\tau(d) = P(D_\tau < d)$. Define X_n as the on-hand inventory level at the beginning of review period n. The (R, s, Q) policy prescribes to order when the inventory level at a review is at or below reorder level s. The amount to order equals Q units, which is delivered after L time units but within the same review period. Since there are never two or more orders outstanding at the same time, this is a simplification compared to the model developed in Section 6.2.4. Hence,

$$X_{n+1} = \begin{cases} (X_n - D_R)^+, & \text{if } X_n > s, \\ \big((X_n - D_L)^+ + Q - D_{R-L}\big)^+, & \text{if } X_n \leq s, \end{cases}$$

where $(A)^+ = \max\{A, 0\}$. The random variable X_{n+1} only depends on X_n and the demand during one review period. Thus $X = \{X_n, n \geq 0\}$ is a homogeneous, one-dimensional Markov chain with state space $\{0, 1, \ldots, s + Q\}$.

In order to define the transition probabilities $P_{ij} = P(X_{n+1} = j | X_n = i)$, we make a distinction between $X_n \leq s$ and $X_n > s$. When $X_n \leq s$,

$$P_{ij} = \begin{cases} \sum\limits_{d=0}^{i-1} g_L(d)(1 - \mathcal{G}^0_{R-L}(i - d + Q)) + (1 - \mathcal{G}^0_L(i))(1 - \mathcal{G}^0_{R-L}(Q)), & j = 0, \\ \sum\limits_{d=0}^{i-1} g_L(d) g_{R-L}(i - d + Q - j) + (1 - \mathcal{G}^0_L(i)) g_{R-L}(Q - j), & 0 < j \leq Q, \\ P(D_R = i + Q - j), & j > Q. \end{cases} \tag{7.1}$$

If $j > Q$ the inventory position can never drop to zero during one review period and therefore P_{ij} is constructed by convolution of D_L and D_{R-L} to D_R. When $X_n > s$,

$$P_{ij} = \begin{cases} 1 - \mathcal{G}_R^0(i), & j = 0, \\ g_R(i-j), & 0 < j \leq i, \\ 0, & j > i. \end{cases} \tag{7.2}$$

This defines the transition matrix P of Markov chain X. The Markov chain X is irreducible and aperiodic since all states communicate. It has a unique stationary distribution π_{IL}, where $\pi_{IL}(j)$ can be interpreted as the limiting probability that the process is in state j at a review moment. The stationary probabilities π_{IL} are given by the solution of

$$\begin{aligned} \pi_{IL}(j) &= \sum_{i=0}^{s+Q} \pi_{IL}(i) P_{ij}, \qquad \text{for } 0 \leq j \leq s+Q \\ \sum_{j=0}^{s+Q} \pi_{IL}(j) &= 1. \end{aligned}$$

We define $\beta(i)$ as the fraction of demand satisfied in a review period, when i units are on hand at the beginning of a review period. As in Chapter 5 and Chapter 6, this equals the complement of the fraction of demand lost. Hence, $\beta(i) = 1 - A(i)$ where for $i \leq s$

$$\begin{aligned} A(i) &= \frac{E\left[\left(D_L - i\right)^+ + \left(D_{R-L} - ((i - D_L)^+ + Q)\right)^+\right]}{E[D_R]} \\ &= \frac{E\left[(D_L - i)^+\right] + \sum_{d=0}^{i-1} g_L(d) E\left[(D_{R-L} - (i - d + Q))^+\right] + \left(1 - \mathcal{G}_L^0(i)\right) E\left[(D_{R-L} - Q)^+\right]}{E[D_R]} \\ &= \frac{\mathcal{G}_L^1(i) + \sum_{d=0}^{i-1} g_L(d) \mathcal{G}_{R-L}^1(i - d + Q) + \left(1 - \mathcal{G}_L^0(i)\right) \mathcal{G}_{R-L}^1(Q)}{E[D_R]}, \end{aligned} \tag{7.3}$$

and for $i > s$

$$A(i) = \frac{E[(D_R - i)^+]}{E[D_R]} = \frac{\mathcal{G}_R^1(i)}{E[D_R]}.$$

The average fill rate is denoted by $\beta = \sum_i \pi_{IL}(i)\beta(i)$. The same analysis can be performed for an (R, s, S) policy. Since the order size in this policy is not fixed, the value of Q has to be replaced by $S - i$ in Equation (7.1) and Equation (7.3).

The objective is to find optimal values of the reorder level and order size for each item while considering the utilization of the storage facility as well as the service level. Since there is a lack of capacity, we set $s + Q = C$ and $S = C$ in the FOSP and POUTP, respectively, to make full use of the available capacity. Consequently, the search space for all possible values of s and Q is one dimensional and restricted. For the general models developed in Section 6.2, this was not the case. Because of the capacity restriction, we cannot derive any convexity properties on the objective function. Consequently, an enumeration procedure for all $C + 1$ values of reorder level s is required to maximize the service level. The available capacity for an item is however also a decision variable. How to determine this value is discussed in the remainder of this section.

7.3.2 Multi-item inventory system

The inventory control problem in hospitals is more complex than the single-item system considered so far. The capacity limitation for each item is part of a larger inventory system with multiple items. Let TC denote the total capacity available for all items, and $TC(k)$ the capacity assigned to item k. Hence, $\sum_k TC(k) \leq TC$. The capacity is, however, expressed in terms of volume, whereas inventory decisions are expressed in terms of number of units. As mentioned in Section 7.2, items are stored in bins. Each bin requires a storage capacity $BC(k)$ for item k and can contain $C(k)$ units. In the multi-item inventory system, the number of bins assigned to each item has to be determined such that the average service level is maximized within the capacity limitation. When we denote this number by $a(k)$ for each item k, then $TC(k) = a(k)BC(k)$ and at most $a(k)C(k)$ units can be stored for item k. The average service level of the entire inventory system is defined as the demand-weighted average service level,

$$\beta_{all} = \sum_k \frac{E[D^k]}{\sum_l E[D^l]} \beta_k,$$

where $E[D^k]$ is the average demand for item k and β_k is the service level for item k. Notice that the service level for item k depends on the replenishment policy and the corresponding values of the inventory control variables. The allocation of the limited storage capacity available, and the determination of the replenishment policy can be solved simultaneously. We propose a knapsack kind of approach for such a solution procedure in which a trade-off has to be made between the increase of the service level for an item versus the allocation of an extra bin to this item. The ratio of this service level increment divided by the extra assigned capacity is computed for each item in every iteration. We assign an extra bin to the item with the highest ratio until all capacity is assigned. To determine the increase in the service level, we determine the optimal control values of the replenishment policy and the corresponding service level as discussed in Section 7.3.1. The solution procedure can be summarized as

Storage capacity allocation

$a(k) = 0$ for all items k
while $(\sum_k TC(k) \leq TC)$
  for each item l
    if $(\sum_k TC(k) + BC(l) \leq TC)$, then
      - determine optimal values of inventory control variables for item l
        where $a(l) = a(l) + 1$
      - determine the new corresponding service level β_l'
    else
      $\beta_l' = 0$
    end if
    $\Delta\beta_l = (\beta_l' - \beta_l)E[D^l] \big/ \sum_k E[D^k]$
    $\Delta C_l = BC(l)$, $\Delta_l = \Delta\beta_l / \Delta C_l$
  end for
  $l^* = \text{argmax}\{\Delta_l\}, a(l^*) = a(l^*) + 1$
end while

Notice that we have developed a same kind of knapsack procedure in Chapter 3. A similar approach can also be used by a hospital manager to decide how much capacity is required and which items should be stored at each of the POU locations.

In the next section, the impact of the different replenishment policies is shown based on a real-world example. It also shows how this inventory model can be used to divide the available capacity among all items.

7.4 Case study

This section illustrates the performance of the most commonly used replenishment policy in hospitals and alternative replenishment policies (i.e., the (R, s, Q) and (R, s, S) policy, respectively) as discussed in Section 7.2 based on the models developed in Section 7.3. It also shows how these models can be used to determine the required storage capacity for each item when a minimal service level constraint should be satisfied.

We observed the (R, s, Q) replenishment policy at the VU University Medical Centre (VUmc) in Amsterdam and at Hospital Amstelland in Amstelveen. Since inventory management is not a main issue in hospitals, there are usually not enough data available on all items stored at each POU location. Most data records concern order quantities and not the actual demand for a certain item. As a result, the available data is very limited. Furthermore, both hospitals have no information on current service levels. Therefore, we can only compare the different models mutually and not with the current situation.

The models described in Section 7.3 can be used for general demand distributions. It is quite common in the literature to assume a Poisson distribution to represent demand in a hospital environment (e.g., Duclos [76], Epstein and Dexter [78]). This is also supported by data from both hospitals. Therefore, we assume that the demand for an item follows a pure Poisson process with an average of λ_τ units over a period of τ time units. We consider the impact of the different replenishment policies by means of an example about infusion liquids at three POU locations (pediatrics, intensive care and obstetrics). The specific parameter values for this product are indicated in Table 7.1 for each of the POU locations. Based on more experiments (not reported here), we conclude that these numbers represent the current practices well for all kinds of items kept on stock in both hospitals.

POU location	L	R	λ_L	λ_R	C
pediatrics	4 hours	1 week	0.114	4.792	5
intensive care	4 hours	3 days	0.924	16.637	30
obstetrics	4 hours	3 days	1.543	27.78	50

Table 7.1: The parameter values corresponding to the current situation for infusion liquid at different POU locations in Hospital Amstelland.

For the (R, s, Q) policy with $s + Q = C$ and the (R, s, S) policy with $S = C$, we computed the fill rate β and average number of review periods between two consecutive orders $OF = 1/\sum_{i=0}^{s} \pi_{IL}(i)$ (i.e., the inverse of the order frequency) for the three POU locations. Both performance measures are illustrated in Figure 7.2 for all possible values of reorder level $s \in \{0, 1, \ldots, C\}$. It is clear from this figure that the service level for the (R, s, Q) policy (denoted by $\beta_{(R,s,Q)}$) is not a convex function in reorder level s for the most common replenishment policy.

Based on enumeration we can determine the optimal value of s for the (R, s, Q) policy. This optimal value is denoted by s^*. For the (R, s, S) policy, the fill rate $\beta_{(R,s,S)}$ increases in s. Consequently, a base-stock policy (i.e., $s = C - 1$) maximizes the service level. However, the order frequency is also the highest for such policies (see Figure 7.2). This results in small and frequent orders which are not desirable for hospitals. Therefore, the value of reorder s is also derived for the (R, s, S) policy with a similar order frequency compared to the best (R, s, Q) policy, denoted by $\bar{s}$. From the results shown in Table 7.2 we conclude that the (R, s, S) policy can improve the service level significantly without an increase of the order frequency.

Based on the results of Table 7.2 and similar research for the VUmc we conclude that with the current (R, s, Q) replenishment policy and the available

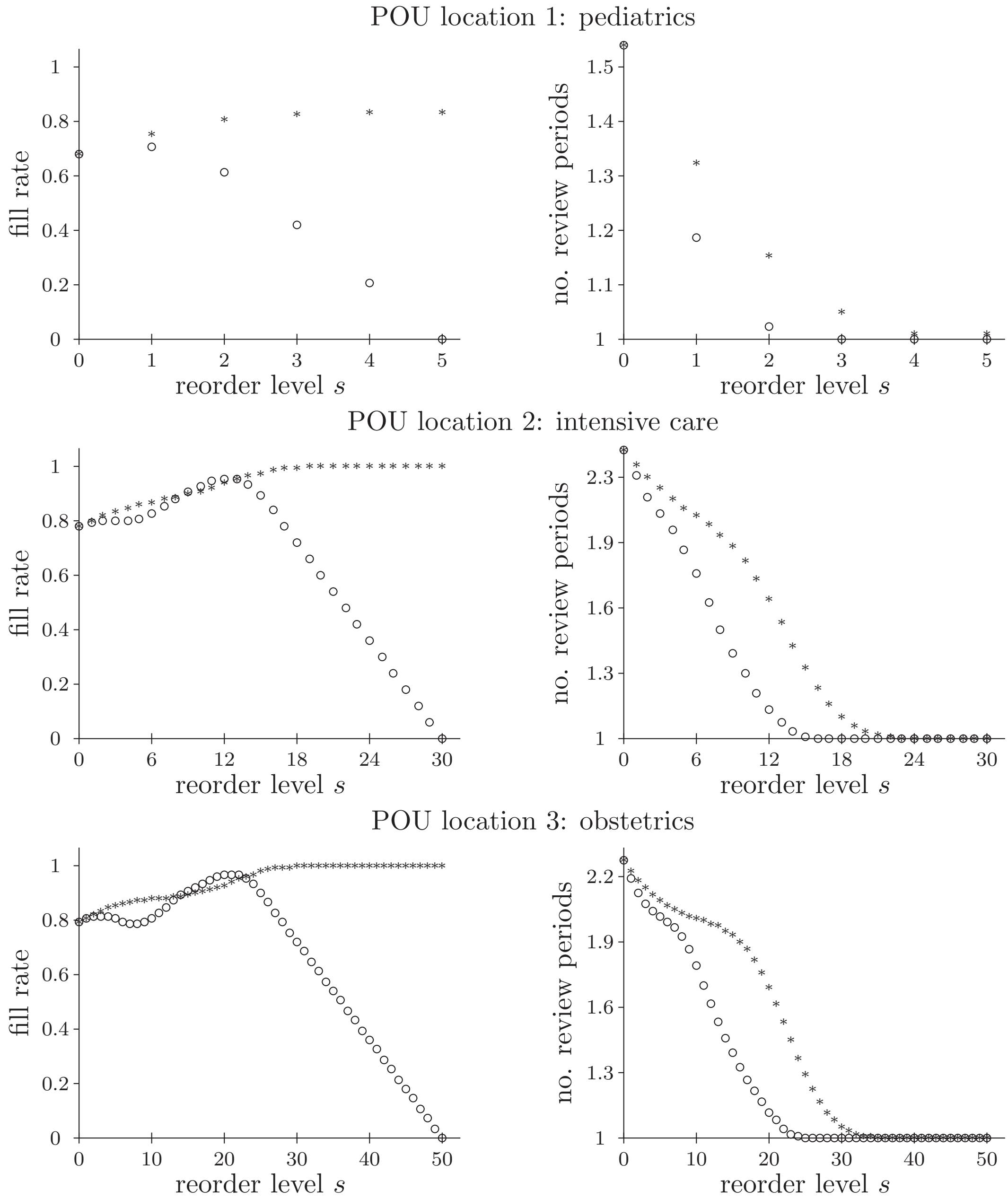


Figure 7.2: The fill rate and the average number of review periods between two subsequent orders for the (R, s, Q) policy (circle) and the (R, s, S) policy (astrics) for three different POU locations.

capacity we can only reach service levels of about 70% to 95%. The capacity C that is assigned to infusion liquid is insufficient and results in service levels that are too low to avoid stock outs and related organizational problems. There are several solutions:

1. use the alternative (R, s, S) replenishment policy with the same order frequency,

2. shorten the replenishment period, or

POU location	s^*	$\beta_{(R,s,Q)}(s^*)$	$OF_{(R,s,Q)}(s^*)$	$\bar{s}$	$\beta_{(R,s,S)}(\bar{s})$	$OF_{(R,s,S)}(\bar{s})$
pediatrics	1	70.42%	1.19	2	80.71%	1.15
intensive care	12	95.27%	1.14	17	99.15%	1.16
obstetrics	21	96.73%	1.08	29	99.65%	1.08

Table 7.2: The fill rate β and order frequency OF for the optimal (R, s, Q) policy (denoted by s^*) and the (R, s, S) policy with a similar order frequency (denoted by $\bar{s}$).

3. increase the available capacity.

The first solution is illustrated in Table 7.2 for the (R, s, S) policy and shows significant service improvements. For a base-stock policy, more smaller orders are placed. This would also happen for shorter review periods. Consequently, the replenishment process would take more time which is not preferred. Therefore, we recommend a restructuring of the available storage capacity. The solution procedure as described in Section 7.3.2 could be used to reallocate items to the limited storage capacity available. For instance, if the service level should at least by 99% for infusion liquid in the (R, s, Q) policy with $s + Q = C$, the following results are obtained: the minimal capacity for pediatrics is 14, for intensive care it is 38 and for obstetrics it becomes 59 units.

The exhaustive search procedure as discussed in Section 7.3.2 can result in large computation times when multiple items are considered at multiple locations. Furthermore, the results are the outcome of a black box for the hospital staff and management. Clearly, a fast and insightful approximation procedure is required to perform an assignment procedure like the one described here. We derive such an efficient inventory rule in the next section.

7.5 Spreadsheet-based inventory rule

The models developed in Section 7.3 and the exhaustive search procedure to find optimal values of the inventory control variables are not insightful and difficult to implement for hospital staff. Especially, when such procedures are repeated iteratively, as indicated in the multi-item model of Section 7.3.2. Therefore, the goal of this section is to develop a heuristic inventory rule for the (R, s, Q) policy that can easily be implemented in a spreadsheet-based program to decide upon the reorder level and order quantity. One of the characteristics of an inventory control system at hospitals is the lack of storage capacity. This limitation is used in the basic thought behind the heuristic. Therefore, such rules are not applicable to the models developed in Chapter 5 and Chapter 6.

The inventory rule consists of several tests. First, we check if the capacity C is sufficient in order to satisfy demand. If the capacity is restrictive, we need to check whether it is likely that this restriction results in out-of-stock occurrences. Therefore, we examine if the reorder level s could be sufficient to be used as safety stock in order to fulfill demand until the next delivery. If this seems to be sufficient, we can determine the value of s such that stock outs are minimized. Otherwise, we need to find a balance between the reorder level and the order quantity.

The capacity is not restrictive when the order quantity is at least the average amount that is asked for during a review period, i.e., $Q \geq \lambda_R$. Another characteristic for this situation is that when no order is placed (i.e., inventory level larger than s) the remaining inventory is sufficient to fulfill the demand until the next possible order delivery (i.e., the demand until the next review and order delivery), or $s + 1 \geq \lambda_R + \lambda_L$. Since $s + Q = C$, the capacity is not restrictive if $C + 1 \geq 2\lambda_R + \lambda_L$. Therefore, we can set $s \in [\lambda_R + \lambda_L - 1; C - \lambda_R]$ to obtain high service levels. We have chosen to set the value of s equal to the middle of this interval.

When there is a shortage of capacity, we want to order at least the average number of units that are asked for during a review period, i.e., $Q = \lambda_R$. This order quantity is on average sufficient to satisfy demand between two order deliveries when orders are placed every review period. Due to the stochastic nature of the demand, we cannot guarantee that an order is placed at each review. Therefore, an approximation is introduced for the probability that orders are placed in two succeeding reviews. This is only likely when $Q \leq \lambda_R$ (i.e., we assume the inventory level to be zero when an order arrives). A new order is placed when the delivered quantity minus the demand between order delivery and the next review moment is less than or equal to s. This is expressed by

$$P(Q - D_{R-L} \leq s) \geq \alpha \quad \Longleftrightarrow \quad P(D_{R-L} \geq Q - s) \geq \alpha. \tag{7.4}$$

The tail probability of a Poisson distribution can be approximated by a normal distribution. Therefore, Equation (7.4) is approximated by

$$1 - \Phi\left(\frac{Q - s - \lambda_{R-L}}{\sqrt{\lambda_{R-L}}}\right) \geq \alpha.$$

Now, we set α sufficiently large such that orders are placed every review period with a high probability ($\alpha = 0.98, \Phi^{-1}(1-\alpha) \approx -2$). When the following inequality is satisfied there is a high probability that an order is placed every review moment

$$\frac{Q - s - \lambda_{R-L}}{\sqrt{\lambda_{R-L}}} \leq -2. \tag{7.5}$$

We can substitute $Q = \lambda_R$ and $s = C - Q$ and check whether Equation (7.5) is satisfied. If it is, these parameter values are most likely to result in a high service level. Otherwise, we have to increase reorder level s (and decrease order quantity Q) until Equation (7.5) is satisfied, i.e.

$$\frac{C - 2s - \lambda_{R-L}}{\sqrt{\lambda_{R-L}}} \leq -2 \quad \Longleftrightarrow \quad s \geq \frac{1}{2}\left\{C - \lambda_{R-L} + 2\sqrt{\lambda_{R-L}}\right\}.$$

The spreadsheet-based inventory rule can be summarized as follows:

1. If $C+1 \geq 2\lambda_R + \lambda_L$, we set s equal to $(C + \lambda_L - 1)/2$ rounded to the nearest integer. Otherwise go to step 2.
2. If $\frac{2\lambda_R - \lambda_{R-L} - C}{\sqrt{\lambda_{R-L}}} \leq -2$, we set s equal to $C - \lambda_R$ rounded to the nearest integer. Otherwise go to step 3.
3. We set s equal to $\frac{1}{2}\left\{C - \lambda_{R-L} + 2\sqrt{\lambda_{R-L}}\right\}$ rounded to the nearest integer.

Notice, when $\lambda_{R-L} = 0$ the second test should be $2\lambda_R \leq C$ since $D_{R-L} = 0$ and Equation (7.4) specifies $Q \leq s$ where $s + Q = C$ and $Q = \lambda_R$.

When we apply this inventory rule to the case study of infusion liquid in Section 7.4, we obtain the following results for reorder level s: for pediatrics 2, for intensive care 13, and for obstetrics 22. The inventory rule can be implemented with the use of a simple spreadsheet program and is, therefore, very appealing to be applied in any hospital. In the following section we illustrate the performance of this inventory rule and the different models of Section 7.3 in a more general setting.

7.6 Numerical results

To goal of this section is to illustrate the performance of the (R, s, Q) and (R, s, S) policy with the characteristics as discussed in Section 7.2, including the inventory rule of Section 7.5. Therefore, we specify test instances in which the average demand in a review period is 5, 15, and 30, while the lead time varies between 0.25 and one times the review period length. In order to test all three situations of the inventory rule, the capacity ranges from λ_R until $3\lambda_R$ with steps of $0.5\lambda_R$. This results in 60 test instances, which cover all three scenarios of the inventory rule.

We have computed the value of the reorder level that maximizes the fill rate in the (R, s, Q) policy with $s + Q = C$ denoted by s^*. As in Section 7.4, $\bar{s}$ represents the value of the reorder level for the (R, s, S) policy where the order frequency is similar as for the (R, s, Q) policy with reorder level s^*. The value of the reorder level based on the heuristic rule of Section 7.5 is denoted by $\hat{s}$. Table 7.3 shows the aggregated results over the four different values of the lead time because they represent a similar capacity limitation. Column 3 represents the average fill rate $\beta_{(R,s,Q)}$ for the (R, s, Q) policy where $s = s^*$. Column 4 and 5 of this table show the service level increase for using the (R, s, S) policy where $s = \bar{s}$ and the service level decrease for the (R, s, Q) policy where $s = \hat{s}$, respectively.

λ_R	C	$\beta_{(R,s,Q)}(s^*)$	$\beta_{(R,s,S)}(\bar{s}) - \beta_{(R,s,Q)}(s^*)$	$\beta_{(R,s,Q)}(s^*) - \beta_{(R,s,Q)}(\hat{s})$
5	5	50.56%	7.72%	10.68%
5	8	72.98%	9.23%	1.15%
5	10	82.37%	8.62%	0.17%
5	13	92.26%	4.95%	0.17%
5	15	96.05%	2.78%	0.35%
15	15	56.32%	9.20%	1.96%
15	23	77.53%	7.92%	0.43%
15	30	88.75%	6.83%	2.02%
15	38	96.04%	3.15%	0.17%
15	45	98.86%	1.03%	0.31%
30	30	59.40%	6.44%	0.19%
30	45	80.23%	6.87%	0.09%
30	60	91.70%	5.14%	2.98%
30	75	97.41%	2.13%	0.31%
30	90	99.52%	0.45%	0.23%

Table 7.3: The average fill rate for using the (R, s, Q) policy with $s = s^*$ and the increase and decrease of the service level for using the (R, s, S) policy with $s = \bar{s}$ and the (R, s, Q) policy with $s = \hat{s}$, respectively. The results are aggregated over $L/R = \{1/4, 1/2, 3/4, 1\}$.

Based on the results in Table 7.3 we conclude that the fill rate can significantly increase when the (R, s, S) policy is applied, while the order frequency remains the same as in the most commonly used (R, s, Q) policy. Furthermore, the inventory rule is very effective to find good values for the reorder level. Only when the capacity is equal to the average demand in a review period for a slow moving item this heuristic rule results in a bad performance. However, such situations should be avoided by hospitals at all times, since the average fill rate is below 60% in those situations. In general we conclude that the inventory rule performs on average within 1% from the optimal reorder level.

7.7 Concluding remarks

The replenishment policy at point-of-use (POU) locations in hospitals can be classified as a lost-sales (R, s, Q) inventory system where the lead time is shorter than the length of a review period. Such inventory systems are commonly applied at hospitals and, in more general settings, in the retail industry because of its simple replenishment process with bar codes. Another characteristic of hospital inventory management is the lack of available storage capacity. We developed a new model for this type of inventory systems and compared it with other replenishment policies to use the available capacity more efficiently and increase the service levels. Order-up-to policies can increase the service levels rapidly. However, they require extra effort for the replenishment process. Therefore, hospitals prefer fixed order sizes and the use of bar codes. We developed a simple spreadsheet-based inventory rule to determine near-optimal values of the reorder levels and order quantities. This inventory rule can easily be embedded in algorithms that assign items to the available capacity at different POU locations (as proposed in Section 7.3.2). It can also be used to determine the required storage capacity. Based on numerical results we show that these conclusions are satisfied for a large set of parameter settings and can therefore be used in more general settings besides hospitals.

Part III

Customer behavior towards stock outs

Chapter 8

Excess demand in practice and literature

The behavior of customers towards out-of-stock occurrences can be different. Either they make a reservation (a backorder model), request an emergency replenishment outside the regular replenishment process (see Chapter 2-3), buy a substitute product or go to a different store (a lost-sales model, see Chapter 4-7). Most inventory systems studied in the literature model this reaction towards excess demand as an assumption. In Section 4.4, we have seen models with a mixture of a lost-sales and backorder assumption. This is however only studied for single-item inventory systems. When customers buy a substitute product at the same retailer or they visit a different retailer which sells the same product brand, the actual demand is not lost. The customer behavior only shifts the demand to a different item or location. Therefore, inventory systems with multiple items and multiple retail locations should be studied to investigate the influence of excess demand. This is the topic of the final part of my thesis. The main goal is to quantify the interaction of the availability of multiple items at multiple locations and the influence it has on inventory control decisions such as order quantities. This is of importance in almost any real inventory system, since no store sells only one item and similar items have some kind of substitution factor (either at an item level or location level). Literature on inventory control problems with different approaches to handle excess demand is discussed in Section 8.1.

In part III we restrict ourselves to situations where customers rent items.

Such systems are good examples to reflect the different types of customer behavior towards excess demand as described above, since most rental items are interchangeable. Our main focus is on libraries, where inventory costs are not relevant but the service perceived by customers is the most important performance measure. The inventory control models developed in this thesis can be applied to other contexts and extended to include inventory costs as well. In Section 8.2 we provide the characteristics of the customer behavior and inventory control systems for rental companies. We also relate these characteristics to logistics problems observed in practical settings regarding inventory control at rental companies. A comparison between current practices at rental companies and the available literature on such inventory systems is presented in Section 8.3. Our contribution is also outlined in that section.

8.1 Literature overview on excess demand

Customers' reaction towards excess demand can be classified in four alternatives: (1) demand is backordered, (2) demand is substituted by a different item at the same location, (3) demand is satisfied by the same item at a different location, (4) demand is lost. Most literature on inventory control systems assumes backordering in case of excess demand. Important references on such systems are, e.g., Federgruen and Zipkin [85], Zheng and Federgruen [326], Porteus [236], Federgruen and Zheng [84], Tijms and Groenevelt [287]. Furthermore, the lost-sales assumption is studied in part II of this thesis. Chapter 4 presents a literature overview on inventory models with this characteristic. The remaining two alternatives to deal with excess demand are discussed in this section. We start with an overview on inventory systems where items can be substituted with other items. The last alternative deals with lateral transshipments. Besides the mixture of lost sales and backorders (see Section 4.4), there is no literature on inventory models in which multiple alternatives to excess demand are included.

Substitute products

One of the first studies on inventory control policies for substitute products is conducted by Ignall and Veinott [125]. Since then, a lot of research has been performed on this topic. We classify these studies based on the number of items included in the model (2 or multiple items) and the type of substitution (full or partial). *Full substitution* means that it is known beforehand which item is the substitute for another item. Moreover, *partial substitution* means that substitution among items is based on proportions or probabilities. Full substitution corresponds to situations in which the probability for item j to be a substitute for item i equals one. This type of substitution is considered in a manufacturer-

items	full substitution	partial substitution
2	Pasternack [233]	Drezner et al. [75], McGillivrary and Silver [198], Parlar and Goyal [232], Wang and Parlar [304], Anupindi et al. [11, 10], Rajaram and Tang [240], Li and Ha [182], Yang and Schrage [317]
n	Ignall and Veinott [125], Bitran [41], Bassok et al. [21], Hsu and Bassok [119], Rao et al. [241], Narayanan and Raman [219]	Smith and Agrawal [273], Van Ryzin and Mahajan [294], Mahajan and Van Ryzin [192, 191], Rajaram and Tory [240], Agrawal and Smith [6], Netessine and Rudi [222]

Table 8.1: An overview of the literature on inventory models with demand substitution for single-period problems.

controlled system, in which a supplier chooses to fill demand for an item with inventory of another item to avoid stock outs. A special type of full substitution is *downward substitution*. This means that demand for item i is met using stock of item j for all $i \geq j$. This is also called *uni-directional substitution* or *one-way substitution*. However, in a retail environment the substitution decision is made by customers. Therefore, partial (or probability) substitutions are common to be found in such situations. This is also referred to as *customer-driven substitution*.

An overview of the literature based on this classification scheme is provided in Table 8.1. In most models unsatisfied demand for an item flows to substitute items, either deterministic or based on deterministic proportions, and otherwise the demand is lost. Li and Ha [182] consider an inventory system with substitute products in which customers can also backorder their demand. Mahajan and Van Ryzin [192, 191] are the only authors to develop a model where the customers demand depends on the available inventory levels.

All papers mentioned so far with stochastic demand and demand substitution consider a single-period inventory problem as extension to the classical newsboy problem (see Chapter 1). Consequently, lead times are not included as well as any reorder level. Therefore, only base-stock policies are considered as replenishment policy. A case study on perishable items in supermarkets by Donselaar et al. [290] shows that such single-period newsboy problems with demand substitution are not sufficient to model all kinds of product groups observed in practical settings. The main reason is that the expiration date is longer than the length of a review period, but also lead times prohibit applications of single-period models. Only daily fresh products like bread can be captured with such models. Multi-period models are developed by Netessine et al. [223], Nagarajan [214], and Yang [317] for settings with customer-driven substitution. All three studies are restricted to models where only 2 items are considered. A continuous review model with unidirectional substitutions is studied by Liu and Lee [185] and Axsäter [16]. The latter two studies are however closely related to inventory systems with lateral transshipment.

Lateral transshipments

Multi-echelon inventory systems are usually used to provide support for products whose customers are distributed over multiple geographical regions. Such systems are characterized by low level stock points (or retailers) that serve as first level of product support to customers, and a depot (or warehouse) that serves as second level of support to retailers. One of the key issues in multiple retailer inventory system is to handle the stock-out position at one retailer when there is inventory available at another retailer. This situation can be resolved either by emergency orders from the higher echelon or it can be allowed to move stock between locations at the same echelon level to enable the sharing of stock. These stock movements are called *lateral transshipments*. The cost of transshipments are in practice generally lower than the shortage cost and the cost of an emergency delivery from upper-level suppliers. Furthermore, the transshipment time is usually shorter than the regular transshipment lead time between the warehouse and retailer. Consequently, lateral transshipments can simultaneously reduce the total system costs and increase the service level at the retailers. Transshipment research is motivated by observations from various industries (e.g., fashion goods, spare parts) and has received a considerable amount of attention in the literature over the past decades.

The earliest models assume instantaneous deliveries (i.e., zero lead times for transshipments and regular deliveries). Consequently, single-period models are proposed for such situations (see also the previous section about substitute products). This assumption is not satisfied in most practical settings. Therefore, we only consider multi-period models in the remainder of this section. For an overview on single-period models with lateral transshipments we refer to Köchel [163]. The complexity of the inventory control problem increases significantly when lead times are included. The information vector that represents the inventory system has to be expended to account for the pipeline inventory. Similar to lost-sales inventory systems, this vector has a length equal to the lead time (see Section 4.1).

There are two types of lateral transshipments: preventive lateral transshipments (PLT) and emergency lateral transshipments (ELT). *Preventive lateral transshipments* reduce the risk of a stock out by redistributing stock between locations to anticipate on future customer demands. Moreover, PLT is called a lateral transshipment for inventory equalization (TIE), since the transshipment decisions are based on the concept of inventory balancing or equalization through stock redistribution. There are three commonly used redistribution rules for PLT (or TIE):

- inventories are redistributed among the retail locations such that all locations satisfy a service level (Reyes and Meade [245]),
- inventories are redistributed among the retail locations to match the ratio of average demand of each retailer to the total demand (Banerjee et al. [18], Burton and Banerjee [48]),
- inventories are redistributed among the retail locations to achieve equal marginal costs over all retailers (Bertrand and Bookbinder [31]).

There are many other possible transshipment policies that can be devised based on the concept of transshipments for inventory equalization.

Emergency lateral transshipments direct an emergency redistribution of items from a retailer with ample stock to a retailer that has reached a stock out. ELT is also called a lateral transshipment based on availability (TBA), since the available stock is transshipped to retailers with no on-hand inventory. Most of the transshipment literature is focused on ELT. Notice the difference between ELT and emergency replenishments. In the former concept, inventory is moved at the same echelon level, whereas in the latter concept inventory is delivered from higher-level stocking points (see also Section 4.5). The available models in the literature for inventory systems with ELT differ on important features that should be taken into account when a model is developed. First, most of the transshipment related research deals with single-item problems where inventory levels are set independently for each individual item (item approach, see part II). An alternative approach is a system approach in which all items are considered when making inventory level decisions (see part I). A system approach is required when there is a capacity limitation or the performance measure is based on all items. A second classification of ELT models is based on the number of echelon levels and retail locations in the supply chain. Single-echelon inventory models assume ample capacity at higher level stocking points, whereas central warehouses can also run out of stock in multi-echelon models. Furthermore, when the number of retail locations is restricted to two retailers it is clear that the other retailer should deliver the emergency transshipment. This is less obvious when lateral transshipments are allowed among more than two retailers. Also the review interval and replenishment policy determine the regular replenishment process in the system. Table 8.2 provides an overview of the literature on models with ELT based on the classifications as described above. Axsäter [16] and Liu and Lee [185] consider uni-direction transshipments, whereas all other papers do not restrict to any transshipment direction. Furthermore, *complete pooling* is usually assumed.

This means that all retailers put their stock available for transshipment, even if this means that they run out of stock. Moreover, excess demand that cannot be satisfied with a lateral transshipment is assumed to be backordered by the central warehouse.

Complete pooling is not assumed by Tagaras and Cohen [282]. They investigate different pooling policies. Therefore, they include an extra inventory control variable to the replenishment policy at each retailer, which represents the inventory level that a retailer wants to maintain when items are transshipped. Such thresholds reduce its own risk of a stock out in the future. The authors conclude that complete pooling dominates partial pooling. They also investigate a second threshold that triggers lateral transshipments, even when there is no shortage. This corresponds to a preventive lateral transshipment policy. Such redistributions of stock have received less attention in the transshipment literature (as described before).

The above mentioned PLT and ELT policies have the disadvantage of not being able to respond to stock outs before or after redistribution of stock. Therefore, Lee et al. [179] propose a third type of lateral transshipment policy, called *service level adjustment*. This policy can reduce the stock-out risk by forecasting stock outs in advance, and efficiently respond to actual stock outs by combining TBA and TIE.

From this literature overview we conclude that generally applicable inventory models have only been developed in case excess demand is assumed to be backordered by customers. Models with a lost-sales assumption on excess demand have been developed in part II. When customers buy (or look for) a substitute product, there are only single-period models available in the literature. This is not a characteristic generally observed in many practical settings, since lead times should be included. The fourth possibility to deal with excess demand is to transship items from a different stocking point at the same echelon level (e.g., retailer or store). This corresponds to situations in which a customer visits another store when excess demand is incurred. Most research on inventory systems with such lateral transshipments focus on emergency transshipments, either for single-period or multi-period models. Models to prevent stock-out occurrences are hardly discussed. In the next section we discuss the importance to model the behavior to excess demand from a practical point of view in a rental environment.

8.2 Rental companies

Nowadays, more items can be rented. One can think of books, movies, cars, but also clothes, toys, equipment and machinery. We also have a much wider class of problems in mind besides these obvious situations in which products are rented. Companies which 'rent out' service personnel, such as technicians, can

	echelon	# items	review	policy	# retailers	lead time
Karmarkar [156]	single	single	P	general	multiple	$L_R = 0, L_E = 0$
Robinson [248]	single	single	P	general	multiple	$L_R = 0, L_E = 0$
Archibald et al. [14]	single	multi	P	general	two	$L_R = 0, L_E = 0$
Tagaras and Cohen [282]	single	single	P	(R, S)	two	$L_R > 0, L_E = 0$
Tagaras [281]	single	single	P	(R, S)	three	$L_R > 0, L_E = 0$
Özdemir et al. [229]	single	single	P	(R, S)	multiple	$L_R = R, L_E = 0$
Lee et al. [179]	single	single	P	(R, S)	multiple	$L_R > 0, L_E = 0$
Hu et al. [121]	two	single	P	(R, s, S)	multiple	$L_R = 0, L_E = 0$
Lee [176]	two	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Axsäter [15]	two	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Dada [65]	two	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Köchel [162]	single	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Ching [59]	two	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E = 0$
Alfredsson and Verrijdt [8]	two	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Grahovac and Chakravarty [97]	two	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Kukreja et al. [171]	single	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E = 0$
Jung et al. [143]	two	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E = 0$
Wong et al. [310]	single	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E = 0$
wong et al. [309]	single	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Kim et al. [161]	single	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E = 0$
Wong et al. [311]	single	multi	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Wong et al. [312]	single	multi	C	$(S-1, S)$	2	$L_R > 0, L_E > 0$
Wong et al. [313]	two	multi	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Kutanoglu [172]	single	single	C	$(S-1, S)$	multiple	$L_R > 0, L_E > 0$
Evers [80]	single	single	C	(s, Q)	multiple	$L_R > 0, L_E = 0$
Needham and Evers [221]	two	single	C	(s, Q)	three	$L_R > 0, L_E = 0$
Evers [81]	single	single	C	(s, Q)	two	$L_R > 0, L_E = 0$
Xu et al. [314]	single	single	C	(s, Q)	two	$L_R > 0, L_E = 0$
Axsäter [17]	single	single	C	(s, Q)	multiple	$L_R > 0, L_E = 0$
Minner et al. [200]	single	single	C	(s, Q)	multiple	$L_R > 0, L_E > 0$
Axsäter [16]	single	single	C	(s, Q)	three	$L_R > 0, L_E = 0$
Ching [58]	single	single	C	(s, Q)	multiple	$L_R = 0, L_E = 0$
Kukreja and Schmidt [170]	single	single	C	(s, S)	multiple	$L_R > 0, L_E = 0$

Table 8.2: An overview of the literature on inventory models with emergency lateral transshipments.

be described equally well with similar models since both classes of problems have the same basic structure. The goal of this section is to introduce and characterize the logistics decision problems in the context of inventory control faced by rental companies, where we focus especially on libraries.

Since items in a rental system are highly interchangeable, there are two types of performance measures used. Either the performance is measured for the single-item (original) demand or for a group of items which includes the substitute products for the original demand. Availability of requested items is one of the most important performance indicators for rental companies to measure effectiveness. Another performance measure for rental companies is the concept of *accessibility*, which measures the amount of time required to obtain an item rather than its immediate availability. In a library setting Kantor [153] defines *availability* as "the extent to which patron needs for specific documents are promptly satisfied". There have been multiple studies of availability in libraries, most of which are based on surveying actual library patrons. An overview is provided by Nisonger [225]. In this thesis we use *title fill rate* as performance measure for libraries. This is defined as the fraction of customer demands for a (book) title that is immediately satisfied. It is a known-item availability measure, contrary to other service level measures such as subject-based fill rates in which the availability of a set of items is measured. Based on an overview study by Nisonger [225], the title fill rate in libraries is on average around 60%. Notice that the title fill rate corresponds to a regular fill rate definition similar to the fill rate used in part II. It can therefore be applied to any inventory control system.

Whenever the original request cannot be satisfied immediately due to the unavailability of an item, we use the following classification for the customer's reaction to handle excess demand:

- The customer wants to have the specific item (e.g., a book title) and makes a reservation. Whenever the same item is available at another location, the unit is immediately taken off the shelves and sent to the customer. Consequently, the customer's request is almost immediately satisfied. If all units of the item are unavailable, the customer has to wait until one of them is returned.
- The customer wants to have a similar item that is immediately available, such that the customer does not leave empty handed.
- The customer leaves the inventory system empty handed.

If an item is taken from a different location in case of excess demand, it is classified as an *emergency lateral transshipment.* Lateral transshipments are used to manage excesses and shortages in multi-location inventory systems. More details on inventory systems with lateral transshipments are discussed in Section 8.1. Making a reservation for an item is seen as a backorder. The second and third option are referred to as substitution and lost sales, respectively. Actually, all options (except the reservation) are a form of lost sales, since the demand for the original item is lost. However, the customer does not perceive the first two options as a lost sale, since (s)he does not leave the system without an item.

Whenever a requested item is available, it is taken by the customer. Typically the time period during which a customer can rent or borrow the item is controlled by the rental company. This time period is referred to as the *rental period*. For books a rental period can be 1 to 3 weeks, whereas movies are usually rented for 1 or 2 days. After such a period the customer can have the option to prolong the rental period. This is especially the case for companies where items are borrowed (like libraries), since customers pay for a subscription instead of for each item borrowed. When the rental period is not extended, the customer returns the item within the rental period to any store which is affiliated with the rental company. For instance, cars can be picked up and dropped off at geographically different locations. The same can occur with books in libraries. The rental company can decide to redistribute the items over the different locations based on its policy and the return patron of customers. This redistribution of items can be based on ownership of the item, but also to prevent future stock outs. Such *preventive lateral transshipments* have been discussed in Section 8.1. Once the item is returned to the store, it is ready to be borrowed or rented again. Especially in a car rental environment the redistribution of items plays an important role. Pachon et al. [230] minimize transportation costs for such a system and formulate a network flow model. Demand is assumed to be deterministic for short-term planning due to reservations. On a larger time scale demand is not known beforehand. Fink and Reiners [89] perform a simulation study to solve the problem with stochastic demands for longer time horizons. A simulation model is also proposed by Barth and Todd [20] on a similar problem where customers share cars. In their model, the waiting time for customers to use a vehicle is analyzed as well as the number of reallocations of vehicles. The transportation of books between multiple locations plays an important role in library environments as well (see Apte and Mason [12]).

The final aspect we would like to mention is the issue of assortment planning and collection control at rental companies. As mentioned before, availability is an important performance measure. Therefore, the rental company needs to have a sufficient number of units available to satisfy the customers' requests. The problem how many units to have in its collection can be solved by making a profile of the current collection and relate this to the usage data of items and to future predictions on customers' requests. The opinion of customers can also be incorporated. Within The Netherlands, such an approach is applied to the collection control in most libraries. It is called the Product-Market-Combination (PMC) model. There are also a few quantitative models that consider the assortment problem for libraries. The method of Dousset and Larbre provides a mathematical approach to balance the connectivity in the collection between genres or topics, and the usage within these parts of the collection. The accessibility of a collection is calculated in the method of Gütersloh based on the average rental period. Both methods are combined in a computer program called 'Collection Optimal' or C-OPT. Even though we focus on models with a service level objective, we would like to mention recent studies by Gerchak et al. [94] and Tang and Deo [284] on assortment planning for rental systems with a cost objective where demand is assumed to be deterministic and stochastic, respectively. Both papers, however, consider a single item at a single location.

Besides availability, collection control is subjected to budget and shelf space constraints on the number of units a rental company can buy, own and display (Sinha and Clelland [270]). There are also other restrictions to which a collection has to satisfy dependent on the type of rental company. For instance, the collection of libraries should be up to date, multiform and representative for the field of knowledge and culture. Due to budget constraints, libraries are more obliged to take effectiveness of book titles into account. However, the quality and multiformity of the collection have to be guaranteed as well. Therefore, libraries have to cooperate and join forces to be market-driven and fulfill a social and cultural task (see also Groeneveld [100]). Such collaborations result in new aspects to take into consideration. For instance, rental companies have to make the trade-off between the extra transportation cost due to the use of items from different locations and the acquisition cost to expand its own collection (Ward et al. [305]). But also ownership becomes an issue when multiple locations share the same item (Henderson [112]).

To our knowledge this is the first research to study the interaction of items stored at multiple locations where all types of customer behavior is observed in case of excess demand: lost sales, backorder, substitution and lateral transshipments. Such characteristics are however seen in many practical applications. Most literature on inventory models only considers one or two of the options. This is also observed in the literature on rental systems. See, for example, the overview by Reisman and Xu [244]. More details on our contribution are discussed in the next section.

8.3 Contribution and outline of part III

The consequences of a stock out depend highly on the situation in which it occurs. For instance, when a certain dairy product is out of stock customers probably take a different brand, whereas for jeans a customer is most likely to visit another store with the same brand. Both alternatives imply different costs and service levels. But also the inventory levels and the corresponding replenishment policies should take this reaction towards excess demand into account. In a lot of practical situations the behavior of customers is unknown. Moreover, each customer behaves differently to a stock-out occurrence. For instance, in rental systems customers can take a different item (without extra expenses) or request for a transshipment from another location. This dynamics should also be reflected in a model to represent the inventory system. To our knowledge, there does not exist a model for such circumstances. Our main contribution in part III is therefore to develop a multi-period model in which (preventive and emergency) lateral transshipments can occur as well as product substitution and a mixture of backorders and lost sales. Each of these alternatives to excess demand occurs according to some probability. In particular we model the customer behavior at libraries. Based on a simulation approach we analyze the performance in terms of (title) fill rate. Consequently, the lead time of lateral transshipments can be neglected. The research closest related to our study is Lee et al. [179] in which both types of lateral transshipments are also included and unsatisfied demand is backordered. We extend this model to include lost sales as well. Moreover, in our model customers can also substitute their original demand when it is out of stock. Consequently, we consider a multi-item inventory system whereas Lee et al. [179] model a single-item inventory system without substitutions. Therefore, a new inventory model is proposed in Chapter 9. The impact of the different alternatives to excess demand on the service level perceived by customers is studied in the next chapter as well.

Chapter 9

Collection control at libraries

As indicated in the previous chapter, customers behave differently towards stock-out occurrences. In such situations they either make a reservation (backorder), request the same item from a different location (emergency lateral transshipment), request a different item from the same location (substitution), or leave the system empty handed (lost sales). On the other hand, companies can anticipate on future stock outs and perform a redistribution of the stock available (preventive lateral transshipment). The main goal of this chapter is to capture this behavior of customers and companies towards excess demand in a model. As explained in Chapter 8, rental companies are the perfect environment to develop such inventory models for. For instance, at libraries it is not clear beforehand how a customer reacts to a stock out. Book titles are highly interchangeable, and a customer might decide to choose another title. However, sometimes a customer wants to read a specific title and makes a reservation. Therefore, all alternative scenarios should be included in an inventory model for libraries.

Libraries can use such models to analyze their collection. They can also use this model to control their collection. The model is especially useful when they want to acquire new book titles to keep the collection up to date and to satisfy customer demands. Libraries have to determine which book titles to acquire and in which quantities. There is however a tight budget available to do such acquisitions since less people are subscribed to a library. To keep customers satisfied and attract more customers, libraries use customer service and fill rates as most important drivers to make decisions regarding their collection. The collection should also maintain a certain level of cultural value. Usually a single item of these (cultural) book titles have to be included in the collection of libraries, and therefore do not need to be included in the model. They however reduce the shelf space available and as a result the value of this parameter needs to be adjusted accordingly. A trade-off has to be made between available budget and capacity, and customer satisfaction to solve the problem of *collection control*. This problem is closely related to a well-known problem in inventory literature, which is called the assortment problem. This is discussed is more detail in Section 9.1. As mentioned before, the service level is used as objective in the collection control problem. Therefore, we develop a simulation procedure in Section 9.2 to analyze the performance of the inventory system at libraries in terms of customer service. The behavior of a customer towards excess demand is expressed as a probability for each possible reaction. The impact of each reaction on the performance of the inventory system is illustrated in Section 9.3. In this way libraries can analyze their current performances. Libraries can also perform such an analysis to determine what the influence is of sharing a collection between libraries on the service perceived by customers. Moreover, they can use the simulation model to solve the problem of collection control. However, simulations require a large amount of computation time. Therefore, an efficient algorithm to solve this problem is discussed in Section 9.5. Our concluding remarks and directions for future research are presented in Section 9.6.

9.1 Assortment problem

One of the most important decisions for libraries is to determine which books to acquire and in which quantities. The objective for libraries is to maximize the customer service within certain budget and space limitations. Besides the acquisition of new books, libraries also have to determine which books to remove from their collection because of the space limitations. We refer to this problem as the *control of the collection*. This problem is closely related to the *assortment problem*, which is a well-known problem in the literature. This problem arises when companies have to determine which items to stock and in which quantities. Therefore, it is mostly observed in a retail environment, where the objective is to minimize costs. The difference between the two problems is that in the assortment problem retailers have to decide which item to include in their assortment where items are sold and not returned, whereas libraries already have a collection which should be controlled and customers return items. In the latter case, products have an (almost) infinite cycle time. Therefore, libraries have to decide whether to acquire new books or to remove books from their collection. This latter decision is not made by retailers in the assortment problem.

The last ten years a couple of interesting papers have appeared on the product assortment problem. However, most of them model the problem as a single-period newsvendor model with substitution (as discussed in Section 8.1). The models differ in their consumer choice model. Van Ryzin and Mahajan [294] study the assortment planning problem with a multinomial logit (MNL) consumer choice model, in which substitutions are based on the choice of assortment. This model is extended by Mahajan and Van Ryzin [193] to include substitutions based on stock-out occurrences as well. Other extensions are proposed Cachon et al. [50, 49], Maddah and Bish [190]. An exogenous demand model is studied by Smith and Agrawal [273] under both types of substitution. Such models are also investigated by Rajaram [239], Agrawal and Smith [6], Kök and Fischer [164], Gaur and Honhon [91], Yücel et al. [320]. The latter papers propose an integer programming formulation to solve the single-location, single-period assortment problem. As compared to the MNL model, the exogenous demand model is more flexible in dealing with the assortment-based and stock-out-based substitutions. However, such models require more effort to characterize the customer's behavior. Therefore, data collection is more important for such models.

In this chapter we extend the beforementioned papers to include lateral transshipments as well as lead times to solve the assignment problem. For libraries it is important to include all alternative customer reactions towards excess demand in the inventory model and, therefore, in the solution approach. The inventory system at libraries is modeled as a multi-period inventory model with multiple locations (see also Section 9.2). A single-period newsvendor model is not sufficient to solve the assortment problem for the extended multi-period model. Therefore, a new solution approach is required. Since the repair kit problem (see Chapter 3) is also a kind of product assortment problem, we propose a similar knapsack procedure to make the trade-off in the assignment problem between the available budget and capacity on the one hand, and the service level on the other hand.

There is a set of library locations L that collaborate to satisfy customer demand. We assume complete pooling, since there is no reason for libraries not to share books (see also Tagaras and Cohen [282]). Each location $l \in L$ has a collection of book titles T_l where the number of units of each title $t \in T_l$ equals S_{lt}. In the assortment problem, the values of S_{lt} have to be determined, where we have capacity and budget restrictions. In our knapsack solution approach, we start with zero stock levels ($S_{lt} = 0$ for all locations l and titles t) and add books to the libraries until the budget is spent or the space limitation is restrictive. Therefore, we consider the increment of the title fill rate when an extra unit of title t is added to the collection at location l (denoted by $\Delta\beta_{title}(S_{lt})$). Furthermore, the price to acquire book title t is denoted by C_t. Consequently, the ratio to determine the impact of one extra unit of title t at location l equals

$$\Delta_{lt} = \frac{\Delta\beta_{title}(S_{lt})}{C_t}.$$

We decided not to include the shelf space in this ratio, since all book titles require the same amount of space. The procedure to solve the assignment problem is the following

ASSORTMENT PROBLEM
$S_{lt} = 0$ for all $l \in L$ and $t \in T$
while(budget and shelf space available)
  - determine the location l^* and title t^* which maximizes Δ_{lt} within
    the budget and space limitations
  - $S_{l^*t^*} = S_{l^*t^*} + 1$
end while

In order to solve the assortment problem we have to compute the increment of the fill rate $\Delta\beta_{title}(S_{lt})$ when an extra title t is acquired at location l. In the next section, we propose a simulation procedure to make such an analysis.

9.2 Simulation model

In this section we propose a simulation model to analyze the performance of inventory systems at libraries. Therefore, we analyze a known collection in the simulation model. The average daily service level perceived by customers is used as performance indicator for the system. Therefore, we use the following definitions to determine the customer satisfaction:

$$\begin{aligned}
\beta_{title} &= \text{title fill rate (immediate availability of original demand)} \\
\beta_{subst} &= \text{fraction of demand substituted} \\
\beta_{ELT} &= \text{fraction of demand satisfied from another location (ELT)} \\
\beta_{back} &= \text{fraction of demand that resulted in reservation (backorder)} \\
\beta_{lost} &= \text{fraction of demand that resulted in a lost sales}
\end{aligned}$$

In our simulation procedure we simulate the customer behavior and analyze the results it has on the inventory levels and customer satisfaction. Customers react differently to a stock out based on the available books on the shelves. Our main goal of this section is to develop a simulation model which represents the customers choice model. Besides customer requests for book titles, the inventory levels also change due to returned books, reservations and the redistribution of books (i.e., preventive lateral transshipments (PLTs)). From a practical point of view we observe the following sequence of activities that occur on a daily basis:

1. demand for book titles occur at different locations and they are taken off the shelves (if possible),

2. customers return books,

3. reserved books are kept aside,

4. the remaining books are redistributed over the multiple locations overnight.

The next day, the same procedure is repeated. The notation and assumptions in each of these steps are discussed in this section.

The number of customers that visit libraries is a stochastic random variable denoted by N and follows a probability distribution $f_N(n)$. The demand of each customer is assumed to be unit sized, i.e., exactly one book title is asked for by each customer. When customers request multiple book titles they are modeled as multiple customers in f_N. With probability g_l a customer visits location l, and wants to rent book title $t \in T_l$ with probability d_{lt}. If the book title is present, it is taken by the customer. Otherwise, there is a probability q that the customer requests a substitutable title. The level in which title t_2 is a substitute for title t is represented by $q_t(t_2)$, where $\sum_{t_2} q_t(t_2) = 1$ and $q_t(t) = 0$. Notice that not necessarily $q_t(t_2) = q_{t_2}(t)$. These probabilities are independent of the collection currently available at the shelves of location l. However, when a substitute title is not available the probability for the other titles to be chosen as substitute increases. Therefore, the probabilities $q_t(t_2)$ should be adjusted to incorporate the current inventory status of all available titles. This adjustment procedure will be explained in more detail later on in this section. If a customer does not want a substitutable title, (s)he visits a location which has ample units of this title on stock with probability s (or the customer requests the book to be transshipped). This represents the probability of an emergency lateral transshipment (see Section 8.1). If the book title is rented out on all locations, the customer makes a reservation according to a probability r (usually $r \leq s$). This corresponds to a backorder. Otherwise the customer leaves the inventory system empty handed. As mentioned before, the inventory status determines the actual behavior of customers. Let us denote the on-hand inventory level of book title t at location l by IL_{lt}. Figure 9.1 gives a summary of the relation between the inventory levels and the customer's reaction towards a stock out.

Based on these probabilities customers rent books for a specific period of time, called the rental period. After this rental period, customers have to return the book or prolong their rental period. Customers can also return the book before the rental period has ended or after the rental period (with a fine of course). The time period between renting the book for the first time and returning the book is called the *total rental period*. The probability distribution of the total rental period for title t is denoted p_t.

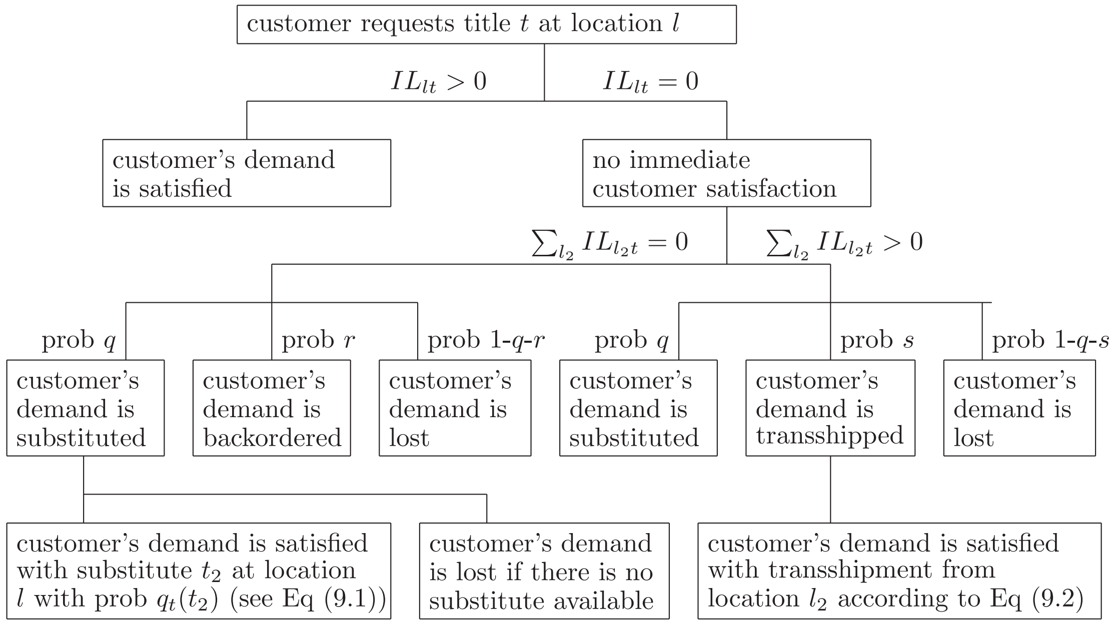


Figure 9.1: The customer's behavior to a stock out depends on the available inventory levels IL_{lt}.

In the simulation procedure we keep track of several variables. Besides the inventory status, we also keep track of all books that are currently rented and will be returned in the future. Furthermore, a list of all reservations is recorded where we assume a first-come first-served priority policy on reserved books. How these variables change during the simulation approach is discussed in the remainder of this section.

Customer demand and behavior

An n-th customer visits location l and requests book title t with probability $P(N \geq n)g_l d_{lt}$. If the title is available (i.e., $IL_{lt} > 0$), the book is taken from the shelves and IL_{lt} decreases by one. Otherwise, the customer has to deal with a stock out. In that case, a substitutable title is looked for with probability q. The probability for a title t_2 to be a substitute is denoted by substitution factor $q_t(t_2)$. However, title t_2 is only a possible substitute for request t if $IL_{lt_2} > 0$. Consequently, the probability for title t_2 to be an actual substitute of title t depends on the availability of all other substitutes t_3 for title t (i.e., $q_t(t_3) > 0$). Therefore, we adjust $q_t(t_2)$ to

$$q'_t(t_2) = \frac{q_t(t_2)}{\sum\limits_{t_3:IL_{lt_3}>0} q_t(t_3)}. \tag{9.1}$$

Notice, when the denominator equals zero, there is no substitute available for title t at location l. Therefore, the customer leaves the system empty handed (see also

Figure 9.1).

Furthermore, in case of a stock out the probability that the customer requests the same title at a different location (i.e., an emergency lateral transshipment) or makes a reservation (i.e., a backorder) also depends on the available inventory. If title t is not available at any location (i.e., $\sum_{l_2} IL_{l_2 t} = 0$), then a reservation is placed for title t with probability r and the customer leaves the system empty handed with probability $1 - q - r$ (excluding the probability to have a lost sales when no substitutable title is available). However, if the same title is available at one of the other locations, customers are more likely to make the reservation or pick up the book themselves because of the direct availability of the book. This is referred to as an *emergency lateral transshipment* (ELT). Consequently, the beforementioned probability r is replaced by s in that case, and the customer's request results in a lost sale with probability $1 - q - s$ (see also Figure 9.1). In case of an ELT, a unit of title t is immediately taken off the shelves at the other location. If multiple locations have a unit available of title t, we choose location l_2 to perform the ELT based on the following criterion:

$$l_2 = \underset{l_3 : IL_{l_3 t} > 0}{\operatorname{argmin}} \left\{ \frac{d_{l_3 t} + \sum_{t_2 : IL_{l_3 t_2} = 0} d_{l_3 t_2} q q'_{t_2}(t)}{IL_{l_3 t}} \right\}. \tag{9.2}$$

Equation (9.2) selects the location for which the total demand for title t per available unit is the lowest. Reducing the inventory level at this location by one unit has (on average) the smallest impact on future stock outs. Therefore, this location l_2 performs the ELT to location l.

Return books

Customers return books during the day. After the demand of all customers has been dealt with, the employees put the returned books back on the shelves. They become available to customers the next day. We assume that customers that have rented a title at a location also return this book to the same location. Notice that a reservation placed at location l is rented at and returned to location l. The same holds for emergency lateral transshipments from location l_2 to location l.

Reservations and redistribution

At the end of each day, the library management has the opportunity to redistribute all available books on the shelves over the various libraries. First, units are kept aside to fulfill reservations. The remaining number of units available for title t is denoted IL_t. To prevent future stock outs as much as possible, these units are

scenario A					
t	1	2	3	4	5
d_{lt}	0.2	0.2	0.2	0.2	0.2

scenario B					
t	1	2	3	4	5
d_{lt}	0.3	0.3	0.15	0.15	0.1

Table 9.1: The demand d_{lt} for title t is the same at all locations l in both scenarios.

assigned to a location l one by one according to the following rule

$$l = \underset{l_2}{\operatorname{argmax}} \left\{ \frac{d_{l_2 t} + \sum\limits_{t_2 : IL_{l_2 t_2} = 0} d_{l_2 t_2} q q'_{t_2}(t)}{IL_{l_2 t} + 1} \right\}. \tag{9.3}$$

Notice the resemblance between Equation (9.2) and Equation (9.3). The former rule selects the location with the lowest (expected) demand per unit, whereas the latter rule selects the location with the highest (expected) future demand per unit (if assigned to this location).

When all available book titles are redistributed the same procedure is repeated to simulate the next day. As mentioned before, we use different definitions to measure the behavior of customers and their satisfaction. During the simulation we keep track of the demand for each title and classify the satisfaction based on the five definitions introduced at the beginning of this section. The impact of this behavior on the customer service is studied in the next section.

9.3 Illustration of excess demand

In this section we analyze the influence of customer behavior towards excess demand on service levels. Therefore, we consider the five types of demand satisfaction introduced in Section 9.2. Two base scenarios are studied, in which customers choose from 5 titles at 3 locations. Each day 5 customers arrive to the inventory system ($N = 5$) and they visit one of the locations according to a uniform distribution (i.e., $g_l = 1/3$). The locations are equivalent to each other and each locations owns 5 units of each book title (i.e., $S_{lt} = 5$). The probability for each title to be requested depends on its popularity, which is presented in Table 9.1 for the two scenarios. In scenario A all book titles have the same popularity, whereas in scenario B titles 1 and 2 are in the same popularity class (most popular) as well as titles 3 and 4 (moderate popular), and title 5 has its own class (least popular). The interchangeability of the titles also depends on the popularity classification. In scenario A $q_{t_1}(t_2) = 0.2$ and Table 9.2 provides the substitution factors in scenario B. The probability distribution for the total rental period in both scenarios is presented in Table 9.3.

		t_2				
		1	2	3	4	5
t_1	1	0	0.5	0.25	0.25	0
	2	0.5	0	0.25	0.25	0
	3	0.25	0.25	0	0.25	0.25
	4	0.25	0.25	0.25	0	0.25
	5	0.1	0.1	0.4	0.4	0

Table 9.2: The substitution factors $q_{t_1}(t_2)$ in scenario B.

	title t	length of total rental period						
		3	6	9	12	18	24	30
A	-	0	0.25	0	0.5	0.25	0	0
B	1	0.25	0.5	0.125	0.125	0	0	0
	2	0.25	0.5	0.125	0.125	0	0	0
	3	0	0	0	0.25	0.5	0.125	0.125
	4	0	0	0	0.25	0.5	0.125	0.125
	5	0	0.25	0	0.5	0.25	0	0

Table 9.3: The probability distribution of the total rental period p_t for title t in both scenarios.

setting	q	s	r	substitution	ELT	PLT	backorder	lost sales
1	0	1	1	no	yes	yes	yes	no
2	0.7	0.3	0.3	yes	yes	yes	yes	no
3	0.7	0.2	0.1	yes	yes	yes	yes	yes
4	1	0	0	yes	no	no	no	no
5	0.5	0.5	0.5	yes	yes	no	yes	no
6	0.5	0.5	0.5	yes	yes	yes	yes	no
7	0.3	0.7	0.7	yes	yes	yes	yes	no
8	1	0	0	yes	no	yes	yes	no

Table 9.4: The different settings to describe the behavior of customers towards stock outs.

For each scenario we want to perform a sensitivity analysis on the behavior of customers. Therefore, we consider 8 different settings according to Table 9.4. In the first setting there are emergency and preventive lateral transshipments as well as backorders (similar to Lee et al. [179]). The substitution of items is added in setting 2 and lost sales are added in setting 3. In setting 4, there are no lateral transshipments. In this setting, customers only substitute demand if possible in case of an out-of-stock occurrence, otherwise the demand is lost. Notice that this setting corresponds to a single location inventory system. Emergency lateral transshipments and backorders are included in setting 5, whereas in setting 6 preventive lateral transshipments are performed. Setting 7 is included to study the impact of the specific values of q, r and s on the satisfied demand. This setting should be compared with setting 2 and setting 6, since they operate under the same customer behavior. In setting 8 customers only substitute their demand in case of excess demand, otherwise it is lost. No emergency lateral transshipments are performed in this setting.

The results for all combinations of the two scenarios and eight settings are provided in Table 9.5. Recall that we used the simulation approach as described in Section 9.2 to derive these numbers, where we simulated 10,000 days. Consequently, the values are not exact but they give a clear indication of the influence of the customer's behavior towards excess demand. When we compare setting 1 to setting 2, we see that substitutions reduce the number of backorders with more than a multiple of 10. Consequently, more books circulate and the title fill rate increases. Notice that about 1% of the customer demand is lost due to the unavailability of a substitutable title. When lost sales are modeled explicitly (setting 3), the title fill rate increases and the number of substitutions decreases. When there are no lateral transshipments (setting 4), the title fill rate is rather low. The title fill rate decreases more when emergency lateral transshipments are allowed

scenario	service definition	setting 1	2	3	4	5	6	7	8
A	β_{title}	56.3%	65.7%	71.0%	60.3%	54.0%	64.6%	61.9%	66.2%
	β_{subst}	0.0%	22.7%	19.7%	34.8%	21.2%	16.9%	11.1%	31.5%
	β_{ELT}	12.7%	7.4%	4.6%	0.0%	16.3%	10.7%	13.2%	0.0%
	β_{back}	31.0%	3.0%	0.6%	0.0%	6.8%	7.0%	13.4%	0.0%
	β_{lost}	0.0%	1.2%	4.2%	4.9%	1.7%	0.8%	0.4%	2.3%
B	β_{title}	61.7%	77.6%	81.4%	71.9%	67.2%	76.5%	74.4%	77.7%
	β_{subst}	0.0%	15.0%	12.6%	25.7%	15.3%	10.9%	7.2%	21.5%
	β_{ELT}	11.0%	4.9%	3.0%	0.0%	12.4%	7.3%	9.3%	0.0%
	β_{back}	27.3%	1.9%	0.4%	0.0%	4.2%	4.7%	8.7%	0.0%
	β_{lost}	0.0%	0.6%	2.6%	2.4%	0.9%	0.5%	0.5%	0.8%

Table 9.5: The results for the different settings and scenarios.

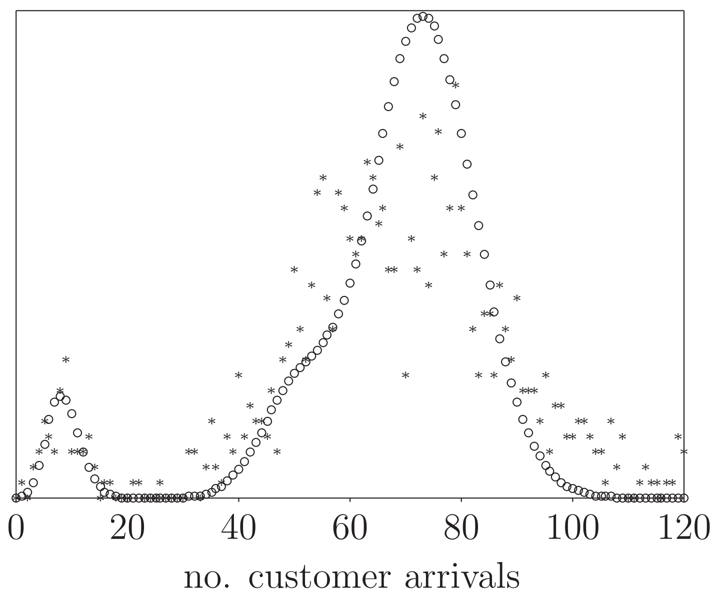
(setting 5). Preventive lateral transshipment increase the service perceived by customers significantly (compare setting 5 with setting 6, and setting 4 with setting 8). Notice that the service level increases more significantly when emergency lateral transshipments are also allowed. Furthermore, the actual values of q, r and s do not have a big influence on the title fill rate, but only on the fraction of demand that results in substitutions, reservations and ELTs, respectively (compare setting 2, 6 and 7).

As mentioned before, the values of q, r and s describe the behavior of customers. They are exogenous variables in most real-world inventory systems. Therefore, no optimization approach can be conducted on these values. However, the results in this section show that all possible reactions to excess demand should be included in the model in order to determine the performance measures of interest for an inventory system. Furthermore, the model is robust to changes in the probabilities for each alternative reaction. How to determine the values of all parameters will be discussed in the next section.

9.4 Case study

In this section we give an illustration how a library can use our model, and how they can use data to derive the necessary values of the input parameters. We use data from the public library in Amstelveen, which has three locations where customers rent books. In particular, we study the data of detective novels over 32 months. There are about 1,200 titles in this category.

First, the arrival process of customers is studied. Figure 9.2 shows a histogram on the number of arrivals on a daily basis. It is common in literature to assume that customers arrive according to a Poisson process (see also Sec-



day	number of customers
Monday	52.01
Tuesday	69.43
Wednesday	76.98
Thursday	76.49
Friday	72.30
Saturday	73.44
Sunday	8.57

Figure 9.2: Histogram of the number of customers that arrive on a daily basis (astrics) and an approximation based on Poisson distributions (circle), where the averages are shown in the table.

tion 7.4). Clearly, such an assumption would not be valid based on our findings of Figure 9.2. However, when we consider each day in the week separately, the arrival process seems to be a Poisson arrival process (see Figure 9.2).

Second, the demand for a specific book of each customer can be derived based on the frequencies that a book has been rented. Therefore, we define M_{lt} as the number of times that book title t has been rented at location l. Consequently,

$$g_l = \frac{\sum_t M_{lt}}{\sum_l \sum_t M_{lt}},$$

and

$$d_{lt} = \frac{M_{lt}}{\sum_t M_{lt}}.$$

Next, the substitution of titles has to be determined. To understand the interchangeability of book titles, we classified the books based on the rent data. More importantly, a distance measure can be computed between two titles (see Boter and Wedel [42]). This measure is used to classify the books. For detective novels, such a classification can result in a class consisting of British detective novels, American detective novels, etc. For each class, we identify all book titles $T_{\mathcal{C}}$ within a class $\mathcal{C}$. Furthermore, the customers are classified similarly, where $B_{\mathcal{C}}$ denotes all customers of class $\mathcal{C}$. For the latter assignment of customers to classes, we use the class which has been rented the most by each customer. About 85% of all customers rent more than half of their books from the same class. The historical data of the rent records of each class give information on the willingness of customers to switch classes and, therefore, on the substitution behavior of customers. So,

$$q_t(t_2) = \frac{N_{\mathcal{C},\mathcal{C}_2}}{\sum\limits_{\mathcal{C}_3} N_{\mathcal{C},\mathcal{C}_3}} \frac{M_{t_2}}{\sum\limits_{\substack{t_3 \in \mathcal{C}_2 \\ t_3 \neq t}} M_{t_3}}, \qquad t_2 \neq t, \tag{9.4}$$

where $M_t = \sum_l M_{lt}$ and $N_{\mathcal{C},\mathcal{C}_2}$ represents the number of times that a customer of class $\mathcal{C}$ rented a book of class $\mathcal{C}_2$. The first term in Equation (9.4) is the probability to switch to class $\mathcal{C}_2$, whereas the second term is the probability to request book title t_2 within the class $\mathcal{C}_2$.

In the final aspect, we analyze the distribution function of the total rental period for each book title. Since each book is rented only a few times a year, there are not much data available on the total rental period of each book title. Mostly, the average rental period is known as well as the standard deviation. A histogram of the total rental period is provided in Figure 9.3. From this figure, we conclude that most of the customers (78.86%) return books according to some random probability function. The remaining customers (21.14%) return the books on a specific day in the week, since Figure 9.3 shows peaks at rental periods that are a multiple of 7. When we decompose the data accordingly, we can use the empirical distribution for the return on the specific days and a negative binomial distribution for the rental period of the majority of the customers. Figure 9.3 shows that this is a good approximation for the distribution of the total rental period.

All values of the input parameters can be used to construct the simulation model of Section 9.2. However, such a model is not appropriate to make decisions on inventory levels because of the excessive computational effort. Therefore, a heuristic model is proposed in the next section.

9.5 Heuristic model

In Section 9.1 we proposed a knapsack heuristic to solve the assortment problem at libraries. Therefore, the increment of the title fill rate has to be computed for each book title t at location l. This is denoted by $\Delta\beta(S_{lt})$. A simulation model is formulated in Section 9.2 to perform such calculations. Consequently, the computational effort to find optimal inventory levels is enormous. Especially when the number of substitutable items is large. Therefore we propose a heuristic procedure to compute $\Delta\beta(S_{lt})$. In such a procedure, we make the following assumptions:

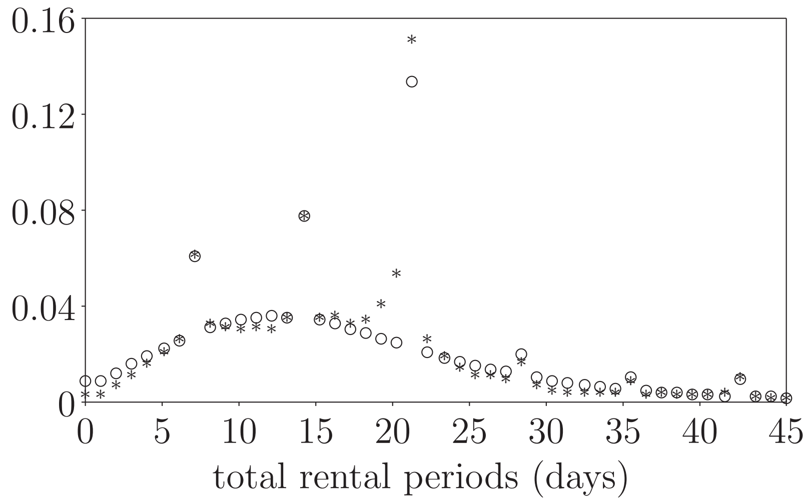


Figure 9.3: The probability distribution function of the rental period (astrics), including an approximation based on a negative binomial distribution (circle).

- customers arrive according to a Poisson process
- the total rental period follows an exponential distribution
- customer leave empty handed when a substitutable title is also out of stock
- customer leave empty handed when the same title is also out of stock at a different location
- there are no reservations allowed
- there are no preventive lateral transshipments allowed

Most of these assumptions are not satisfied in real-world inventory systems. However, they can be used in the procedure to solve the assortment problem and determine the relative increase of the service level. Based on these assumptions, we model the inventory system as a network of servers with no queueing buffer. We have represented this approximation model in Figure 9.4. Each server corresponds to a book title. The number of servers related to book title t at location l is denoted S_{lt} (as introduced in Section 9.1). Customers arrive to one of the S_{lt} servers according to a Poisson process with mean $\lambda_{lt} = E[N]g_l d_{lt}$, where $E[N]$ is the expected number of customers to arrive on a day ($E[N] = \sum_n n f_N(n)$). When there is a server available, this corresponds to an available book on the shelves. As a result, the customer takes the book from the shelves and the book is returned after a total rental period. In the mean time, the server is busy. This time period follows an exponential distribution with rate $\mu_t = 1/E[P]$ where $E[P] = \sum_i i p_t(i)$. When all S_{lt} servers are busy, this means that all units of title t are rented to customers at location l. A new customer demand results in excess demand. Since there is no queueing buffer for the set of servers, the customer's

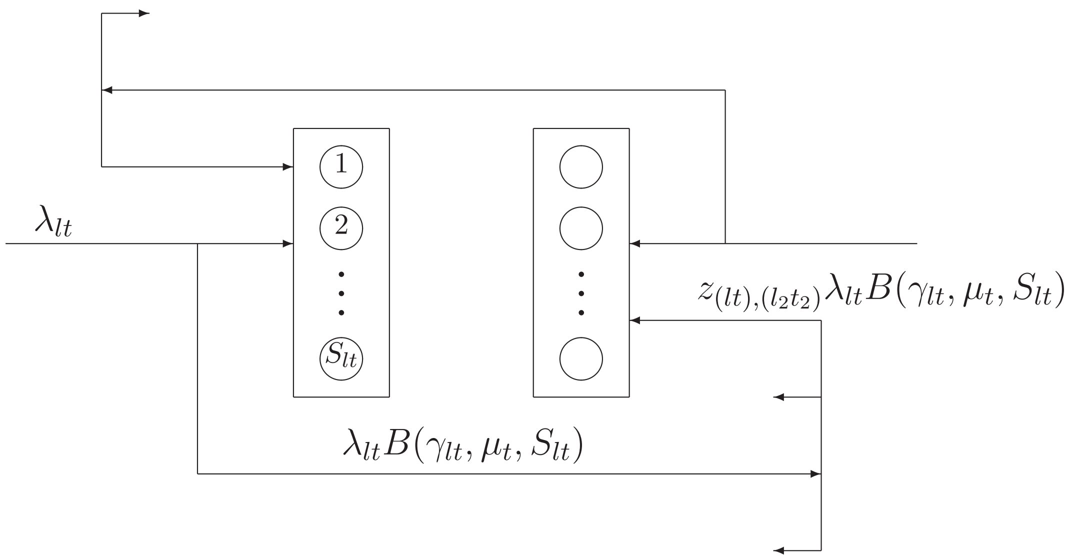


Figure 9.4: Network of servers to represent the customer behavior for book title t at location l.

demand has to be redirected to either a substitutable title $t_2 \neq t$ or to an alternative location $l_2 \neq l$ (see also Figure 9.1). We denote the probability for the demand of a single customer for title t at location l to be redirected to title t_2 at location l_2 by $z_{(lt),(l_2t_2)}$.

As discussed in Section 9.2, this alternative behavior in case of a stock out depends on the availability of books for each alternative. Therefore, we denote the probability that a request for title t at location l cannot be satisfied by $B(\gamma_{lt}, \mu_t, S_{lt})$, where γ_{lt} is the effective arrival rate of demand for title t at location l. This effective arrival rate consists of the original demand λ_{lt} and the redirected requests. Hence,

$$\gamma_{lt} = \lambda_{lt} + \sum_{l_2}\sum_{t_2} z_{(l_2t_2),(lt)}\lambda_{l_2t_2}B(\gamma_{l_2t_2}, \mu_{t_2}, S_{l_2t_2}). \tag{9.5}$$

This system corresponds to a network of Erlang-loss systems, as indicated by Figure 9.4. Such systems have been studied in queueing theory. As mentioned by Riordan [246], the redirected interflow time periods follows a hyperexponential distribution. Consequently, the analysis of the performance for such systems is difficult. However, Koole and Talim [165] approximate the performance with a similar queueing network where they assume that the redirected requests occur according to a Poisson process. We use the same assumption to simplify the calculations. This is also justified, since we do not approximate the actual service perceived by customers in this heuristic model but only the relative increase of the service level when the inventory levels change is of our interest. Consequently,

$$B(\gamma_{lt}, \mu_t, S_{lt}) \approx \frac{\rho^{S_{lt}}}{S_{lt}!\left(1 + \rho + \cdots + \frac{\rho^{S_{lt}}}{S_{lt}!}\right)}, \tag{9.6}$$

where $\rho = \gamma_{lt}/\mu_{lt}$. When this expression is substituted in Equation (9.5) we have to solve a set of equations to determine γ_{lt}. Koole and Talim [165] propose an iterative procedure to solve this problem. We use a similar procedure, where

$$\gamma_{lt}^{(n)} = \begin{cases} \lambda_{lt} + \sum_{l_2} \sum_{t_2} z_{(l_2 t_2),(lt)} \lambda_{l_2 t_2} B(\gamma_{l_2 t_2}^{(n-1)}, \mu_{t_2}, S_{l_2 t_2}), & \text{if } n \geq 1, \\ \lambda_{lt}, & \text{if } n = 0. \end{cases} \tag{9.7}$$

These effective arrival rates determine the availability of a book title and, therefore, they also influence the customer's behavior towards a stock out of title t at location l. Next, the probabilities $z_{(lt),(l_2 t_2)}$ of the customer's behavior in case of a stock out have to be derived.

First, excess demand for title t is substituted by title t_2 with probability $z_{(lt),(lt_2)}$ at location l, where

$$z_{(lt),(lt_2)} = q \frac{q_t(t_2)\big[1 - B(\gamma_{lt_2}, \mu_{t_2}, S_{lt_2})\big]}{\sum_{t_3 \neq t} q_t(t_3)\big[1 - B(\gamma_{lt_3}, \mu_{t_3}, S_{lt_3})\big]}, \qquad t \neq t_2. \tag{9.8}$$

Second, an ELT from location l_2 is performed with probability $z_{(lt),(l_2 t)}$ when title t is out of stock at location l, where

$$z_{(lt),(l_2 t)} = s \frac{1 - B(\gamma_{l_2 t}, \mu_t, S_{l_2 t})}{\sum_{l_3 \neq l} \big[1 - B(\gamma_{l_3 t}, \mu_t, S_{l_3 t})\big]}, \qquad l \neq l_2. \tag{9.9}$$

This defines the network of book titles and the flow of customer demand. The title fill rate is now expressed as

$$\beta_{titile} = \sum_l \sum_t g_l d_{lt} \big[1 - B(\gamma_{lt}, \mu_t, S_{lt})\big].$$

This expression can be used to compute Δ_{lt} in the solution approach for the assortment problem.

To illustrate the performance of this heuristic model, let us consider the example of scenario B and setting 3 from Section 9.3 where $C_{lt} = 1$ for all locations and book titles. Furthermore, we restrict the capacity to 75 book titles (as in Section 9.3). When we apply the solution algorithm, it results in the assortment of Table 9.6.

Based on this example, we illustrate that the heuristic model can be used to compare inventory systems with different stock levels. Such comparisons are useful when decisions have to be made regarding inventory levels (e.g., in the assortment problem). Notice that the heuristic model does not approximate the actual performance of the inventory system, and can therefore only be used to make relative decisions.

		title t				
		1	2	3	4	5
location l	1	6	6	5	5	3
	2	6	6	5	5	3
	3	6	6	5	5	3

customer satisfaction	
β_{title}	84.85%
β_{sub}	10.34%
β_{ELT}	2.30%
β_{back}	0.37%
β_{lost}	2.15%

Table 9.6: The solution to the assortment problem, including the resulting customer service for scenario B.

9.6 Concluding remarks

In this chapter, we illustrated that inventory models should include all types of customer behavior towards a stock-out occurrence to accurately represent real inventory systems. When substitutions and lateral transshipments are observed, the inventory system should be modeled as a multi-period inventory model with multiple locations. A simulation model is required to analyze such inventory systems due to the behavior of customers. For instance, the reaction towards a stock out depends on the availability of alternative items at alternative locations. Such relations and dependencies result in a complex model. Simulation models can be used to perform calculations for such inventory systems. We illustrated in Section 9.3 how a simulation model can be used to perform sensitivity analyses. However, when the simulation model is used in an optimization problem (such as the assortment problem), the computation times increase rapidly. Therefore, we proposed a heuristic model which can be used to help decide on inventory levels. In particular, we focused on collection control at libraries and the availability of books.

Chapter 10

Conclusions

In this chapter we look back on the contents of this thesis and make our final statements.

Customers are the most important aspect of supply chains. It is important to have a customer-focused inventory system and have a high availability of items to satisfy customer requests. However, most models in inventory theory are focused on costs. Another gap between inventory literature and real-world applications is the modeling of customer behavior towards stock-out occurrences. In most models in inventory theory, excess demand is assumed to be backordered. However, in practice customers can choose between different alternatives (substitute products, a different store, etc.). In this thesis, we have bridged the gap between literature and practice with respect to both aspects. Since most inventory systems have periodic order moments in practice, we only studied periodic review models.

In part I, we considered an after-sales service where customers' requests consist of multiple items. When one or more units are not sufficiently available, the entire request is cancelled and an emergency order is placed outside the regular replenishment process. Consequently, the inventory levels remain the same. This dependency between customer satisfaction and availability is hardly studied in the literature. Most inventory models assume an item approach, in which the customer satisfaction is defined in terms of availability of only one item. However, when multiple items are requested by one customer, (s)he is only satisfied when the entire request is fulfilled. This is the first study in which customers do not accept partially fulfilled requests. In Chapter 3, we proposed a solution procedure to determine (near) optimal stock levels. Moreover, we performed a case study in which we show that the solution procedure can improve current practices significantly (31% service level improvement against current holding costs). The model could easily be extended to include budget constraints or storage capacity limitations.

In part II, a single-item inventory system is investigated where excess demand is assumed to be lost instead of backordered. From the literature overview on lost-sales inventory systems, we concluded that not much is known about an optimal replenishment policy when excess demand is lost. The properties and numerical results that have been derived for the optimal order quantities show that there is no structure for an easy-to-understand optimal replenishment policy which can be implemented in real-life applications. However, the most effective approximation policies that have been proposed in the literature include some kind of delay in the ordered quantities. For continuous review policies this is explicitly included, whereas for periodic review systems it is implicitly included with a maximum order size. This delay prevents too many orders to arrive after a period with many customer demands. In a backorder setting these items are already allocated to excess demand, whereas in a lost-sales setting this demand is lost. Consequently, in a lost-sales setting the order sizes do not have to raise the inventory position as much as in a backorder setting when the on-hand inventory level is low.

We developed new models for such new type of policies for periodic review inventory systems. We have considered systems without and with fixed order costs, and we have relaxed assumptions regarding the lead time. We compared the performance to well-known replenishment policies such as order-up-to policies and fixed order size policies. Numerical experiments show that such policies with delay result in near-optimal costs for periodic review models. There is no such comparison for continuous review systems. This is an interesting aspect to investigate in the future.

The difficulty with exact models in a lost-sales context is the computational complexity. Therefore, we also proposed an approximation model in which the steady-state behavior of the inventory system is approximated. Such approximation procedures are very useful to determine near-optimal order quantities. The performance of such procedures is illustrated for models with a cost objective and when a service level constraint is included next to a cost objective. In particular, we considered a case study in which the inventory levels at hospitals have to be determined. Because of specific characteristics, we derived a simple and efficient inventory rule to make near-optimal replenishment decisions. The heuristic rule can easily be implemented in a spreadsheet-based program, and is therefore very appealing to be used in practice.

An interesting direction for future research is the influence of the beforementioned policies with delay on the entire supply chain. When the ordering process is smoothened over time, the variability in the demand for production and transportation also reduces. There is hardly any research performed on the impact of lost-sales in a multi-echelon setting.

In part III, we do not assume to know the customer behavior towards stock outs. In case of excess demand a customer can either make a reservation (i.e., backorder), look for substitute products, go to another store or the demand is lost. We also included lateral transshipments to prevent stock outs. Since the behavior of customers towards stock-out occurrences depend on the inventory levels of the available items, it becomes rather complex to develop an exact mathematical model. Therefore, we proposed a simulation model to analyze such inventory systems. However, simulation is not recommended when optimization decisions have to be made on the inventory levels for such systems. Therefore, a heuristic procedure is proposed to approximate the performance of the inventory system. The assortment problem is used to illustrate the performance of this heuristic model.

Based on this study we conclude that models become rather complex when to model the customer behavior towards stock-out occurrences besides a backorder assumption. However, it is necessary to include this behavior to guarantee high service levels and keep customers satisfied. In order to use any of the models in practice, we recommend the usage of our approximation procedures. Such procedures result in near-optimal stock levels without large computation times.

Bibliography

[1] Uitgeverijen op de Amazon-toer. `http://www.marketingonline.nl/nieuws/index2005-44.html`, 2005.

[2] Wholesale and retail trade. `http://www.census.gov/prod/2008pubs/09statab/domtrade.pdf`, 2007.

[3] K. Aardal, O. Jonsson, and H. Jonsson. Optimal inventory policies with service-level constraints. *Journal of the Operational Research Society*, 40:65–73, 1989.

[4] P.L. Abad. Optimal price and order size under partial backordering incorporating shortage, backorder and lost sale costs. *International Journal of Production Economics*, 114:179–186, 2008.

[5] N. Agrawal and M.A. Cohen. Optimal material control in an assembly system with component commonality. *Naval Research Logistics*, 48:409–429, 2001.

[6] N. Agrawal and S. Smith. Optimal retail assortments for substitutable items purchased in sets. *Naval Research Logistics*, 50:793–822, 2003.

[7] N. Agrawal and S.A. Smith. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics*, 43:839–861, 1996.

[8] P. Alfredsson and J. Verrijdt. Modeling emergency supply flexibility in a two-echelon inventory system. *Management Science*, 46:1416–1431, 1999.

[9] J. Andersson and P. Melchiors. A two-echelon inventory model with lost sales. *International Journal of Production Economics*, 69:307–315, 2001.

[10] R. Anupindi and Y. Bassok. Centralization of stocks: Retailers vs. manufacturer. *Management Science*, 45:178–191, 1999.

[11] R. Anupindi, M. Dada, and S. Gupta. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science*, 17:406–423, 1998.

[12] U.M. Apte and F.M. Mason. Analysis and improvement of delivery operations at the San Francisco Public Library. *Journal of Operations Management*, 24:325–346, 2006.

[13] B.C. Archibald. Continuous review (s, S) policies with lost sales. *Management Science*, 27:1171–1177, 1981.

[14] T. W. Archibald, S. A. E. Sassen, and L. C. Thomas. An optimal policy for a two depot inventory problem with stock transfer. *Management Science*, 43:173–183, 1997.

[15] S. Axsäter. Modeling emergency lateral transshipments in inventory systems. *Management Science*, 36:1329–1338, 1990.

[16] S. Axsäter. Evaluation of unidirectional lateral transshipments and substitutions in inventory systems. *European Journal of Operational Research*, 149:438–447, 2003.

[17] S. Axsäter. A new decision rule for lateral transshipments in inventory systems. *Management Science*, 49:1168–1179, 2003.

[18] A. Banerjee, J. Burton, and S. Banerjee. A simulation study of lateral transshipments in single supplier, multiple buyers supply networks. *International Journal of Production Economics*, 81–82:103–114, 2003.

[19] S.K. Bar-lev, M. Parlar, and D. Perry. Impulse control of a Brownian inventory system with supplier uncertainty. *Stochastic Analysis and Applications*, 11:11–27, 1993.

[20] M. Barth and M. Todd. Simulation model performance analysis of a multiple station shared vehicle system. *Transportation Research*, Part C 7:237–259, 1999.

[21] Y. Bassok, R. Anupindi, and R. Akella. Single-period multiproduct inventory models with substitution. *Operations Research*, 47:632–642, 1999.

[22] M.J. Beckmann and S.K. Srinivasan. An (s, S) inventory system with Poisson demands and exponential lead time. *OR Spektrum*, 9:213–217, 1987.

[23] P.C. Bell. A new procedure for the distribution of periodicals. *Journal of the Operational Research Society*, 29:427–434, 1978.

[24] P.C. Bell. Adaptive sales forecasting with many stockouts. *Journal of the Operational Research Society*, 32:865–873, 1981.

[25] P.C. Bell. Forecasting demand variation when there are stockouts. *Journal of the Operational Research Society*, 51:358–363, 2000.

[26] R. Bellman, I. Glicksberg, and O. Gross. On the optimal inventory equation. *Management Science*, 2:83–104, 1955.

[27] A. Bensoussan, M. Cakanyildirum, J.A. Minjarez-Sosa, A. Royal, and S.P. Sethi. Inventory problems with partially observed demands and lost sales. *Journal of Optimization Theory and Applications*, 136:321–340, 2008.

[28] A. Bensoussan, M. Crouhy, and J. Proth. *Mathematical Theory of Production Planning*. North-Holland, Amsterdam, 1983.

[29] E. Berk and A. Arreola-Risa. Note on 'future uncertainty in EOQ models'. *Naval Research Logistics*, 41:129–132, 1994.

[30] E. Berk, Ü. Gürler, and R.A. Levine. Bayesian demand updating in the lost sales newsvendor problem: A two-moment approximation. *European Journal of Operational Research*, 182:256–281, 2007.

[31] L.P. Bertrand and J.H. Bookbinder. Stock redistribution in two-echelon logistics systems. *Journal of the Operational Research Society*, 49:966–975, 1998.

[32] H.K. Bhargava, D. Sun, and S.H. Xu. Stockout compensation: Joint inventory and price optimization in electronic retailing. *INFORMS Journal on Computing*, 18:255–266, 2006.

[33] U.N. Bhat. *Elements of Applied Stochastic Processes.* Wiley, New York, 1984.

[34] M. Bijvank. Car stock management - the need to close the gap between theory and practice. In D. Kisperska-Moon, editor, *Tenth ELA Doctorate Workshop*, 2005.

[35] M. Bijvank, S. Bhulai, and W.T. Huh. Replenishment policies for lost-sales inventory systems with periodic reviews and order cost. Working paper to be submitted for publication, 2009.

[36] M. Bijvank and S.G. Johansen. Periodic-review lost-sales inventory models with constant lead times of any length. Submitted for publication, 2009.

[37] M. Bijvank and I.F.A. Vis. Quality-driven inventory management at hospitals. Submitted for publication, 2009.

[38] M. Bijvank and I.F.A. Vis. A retail inventory system with lost sales and a service level criterion. Submitted for publication, 2009.

[39] M. Bijvank, I.F.A. Vis, and J. Boter. Determination of assortment at rental companies. Working paper to be submitted for publication, 2009.

[40] M. Bijvank, I.F.A. Vis, and G. Koole. Optimizing a general repair kit problem with a service constraint. Submitted for publication, 2009.

[41] G. Bitran and S. Dasu. Ordering policies in an environment of stochastic yields and substitutable demands. *Operations Research*, 40:1999–1017, 1992.

[42] J. Boter and M. Wedel. User categorization of public library collections. *Library & Information Science Research*, 27:190–202, 2005.

[43] T. Boyaci and G. Gallego. Supply chain coordination in a market with customer service competition. *Production and Operations Management*, 13:3–22, 2004.

[44] D.J. Braden and M. Freimer. Information dynamics of censored observations. *Management Science*, 37:1390–1404, 1982.

[45] M.L. Brandeau, F. Sainfort, and W.P. Pierskalla. *Operations research and health care - A handbook of methods and applications.* Kluwer Academic Publishers, 2004.

[46] S. Brumelle and D. Granot. The repair kit problem revisited. *Operations Research*, 41:994–1006, 1993.

[47] D.J. Buchanan and R.F. Love. A (Q, R) inventory model with lost sales and Erlang-distributed lead times. *Naval Research Logistics Quarterly*, 32:605–611, 1985.

[48] J. Burton and A. Banerjee. Cost-parametric analysis of lateral transshipment policies in two-echelon supply chains. *International Journal of Production Economics*, 93–94:169–178, 2005.

[49] G.P. Cachon and A.G. Kök. Category management and coordination in retail assortment planning in the presence of basket shopping consumers. *Management Science*, 53:934–951, 2007.

[50] G.P. Cachon, C. Terwiesch, and Y. Xu. Retail assortment planning in the presence of consumer search. *Manufacturing & Service Operations Management*, 7:330–346, 2005.

[51] C.-T. Chang and T.Y. Lo. On the inventory model with continuous and discrete lead time, backorders and lost sales. *Applied Mathematical Modelling*, 33:2196–2206, 2009.

[52] Y.F. Chen, S. Ray, and Y. Song. Optimal pricing an inventory control policy in periodic-review systems with fixed ordering cost and lost sales. *Naval Research Logistics*, 53:117–136, 2006.

[53] F. Cheng, M. Ettl, G.Y. Lin, and D.D. Yao. Inventory-service optimization in configure-to-order systems. *Manufacturing & Service Operations Management*, 4, 2002.

[54] F. Cheng and S.P. Sethi. Optimality of state-dependent (s, S) policies in inventory models with Markov-modulated demand and lost sales. *Production and Operations Management*, 8:183–192, 1999.

[55] K.L. Cheung and W.H. Hausman. Multiple failures in a multi-item spare inventory model. *IIE Transactions*, 27:171–180, 1995.

[56] C. Chiang. Optimal ordering policies for periodic-review systems with replenishment cycles. *European Journal of Operational Research*, 170:44–56, 2006.

[57] C. Chiang. Optimal ordering policies for periodic-review systems with a refined intra-cycle time scale. *European Journal of Operational Research*, 177:872–881, 2007.

[58] W.-K. Ching, W.-O. Yuen, and A.W. Loh. An inventory model with returns and lateral transshipments. *Journal of the Operational Research Society*, 54:636–641, 2003.

[59] W.K. Ching. Markov-modulated Poisson processes for multi-location inventory problems. *International Journal of Production Economics*, 53:217–223, 1997.

[60] C.W. Chu, B.E. Patuwo, A. Mehrez, and G. Rabinowitz. A dynamic two-segment partial backorder control of (r, Q) inventory system. *Computers & Operations Research*, 28:935–953, 2001.

[61] A.J. Clark and H. Scarf. Optimal policies for a multi-echelon inventory problem. *Management Science*, 5:475–490, 1960.

[62] M.A. Cohen, P.R. Kleindorfer, and H.L. Lee. Service constrained (s, S) inventory system with priority demand classes and lost sales. *Management Science*, 34:482–499, 1988.

[63] M.A. Cohen, P.R. Kleindorfer, and H.L. Lee. Near-optimal service constrained stocking policies for spare parts. *Operations Research*, 37:104–117, 1989.

[64] S.A. Conrad. Sales data and the estimation of demand. *Operations Research Quarterly*, 27:123–127, 1976.

[65] M. Dada. A two-echelon inventory system with priority shipments. *Management Science*, 38:1140–1153, 1992.

[66] K. Danas, P. Ketkidis, and A. Roudsari. A virtual hospital pharmacy inventory: An approach to support unexpected demand. *International Journal of Medical Marketing*, 2:125–129, 2002.

[67] C Das. The $(S - 1, S)$ inventory model under time limit on backorders. *Operations Research*, 25:835–850, 1977.

[68] P.J. Daugherty, J.J. Kasulis, and R.G. Richey. The role of customer service in trade promotion decisions. *International Journal of Logistics Management*, 13:43–56, 2002.

[69] S. Dayanik, J.-S. Song, and S.H. Xu. The effectiveness of several performance bounds for capacitated production, partial-order-service, assemble-to-order systems. *Manufacturing & Service Operations Management*, 5:230–251, 2003.

[70] A.G. De Kok. Approximations for a lost-sales production/inventory control model with service level constraints. *Management Science*, 31:729–737, 1985.

[71] R. Dekker, R.M. Hill, M.J. Kleijn, and R.H. Teunter. On the $(S - 1, S)$ lost sales inventory model with priority demand classes. *Naval Research Logistics*, 49:593–610, 2002.

[72] N. Dellaert and E. van de Poel. Global inventory control in an academic hospital. *International Journal of Production Economics*, 46–47:277–284, 1996.

[73] B.T. Doshi, F.A. Van der Duyn Schouten, and A.J.J. Talman. A production-inventory control model with a mixture of back-orders and lost-sales. *Management Science*, 24:1078–1086, 1978.

[74] B. Downs, R. Metters, and J. Semple. Managing inventory with multiple products, lags in delivery, resource constraints, and lost sales: A mathematical programming approach. *Management Science*, 47:464–479, 2001.

[75] Z. Drezner, H. Gurnani, and B.A. Pasternack. An EOQ model with substitutions between products. *Journal of the Operational Research Society*, 46:887–891, 1995.

[76] L.K. Duclos. Hospital inventory management for emergency demand. *International Journal of Purchasing and Materials Management*, 29:30–37, 1993.

[77] G.D. Eppen and L. Schrage. Centralized ordering policies in a multi-warehouse system with leadtimes and random demand. In L.B. Schwarz, editor, *Multi-Level Production/Inventory Control Systems: Theory and Practice*, pages 51–69. North Holland Amsterdam, 1981.

[78] R.H. Epstein and F. Dexter. Economic analysis of linking operating room scheduling and hospital material management information systems for just-in-time inventory control. *International Anesthesia Research Society*, 91:337–343, 2000.

[79] T.R. Ervolina, M. Ettl, Y.M. Lee, and D.J. Peters. Managing product availability in an assemble-to-order supply chain with multiple customer segments. *OR Spectrum*, 31:257–280, 2009.

[80] P.T. Evers. Hidden benefits of emergency transshipments. *Journal of Business Logistics*, 18:55–76, 1997.

[81] P.T. Evers. Heuristics for assessing emergency transshipments. *European Journal of Operational Research*, 129:311–316, 2001.

[82] N. Fabbe-Costes and M. Jahre. Supply chain integration and performance: A review of the evidence. *The International Journal of Logistics Management*, 29:130–154, 2008.

[83] A. Federgruen. Centralized planning models for multi-echelon inventory systems under uncertainty. In S.C. et al Graves, editor, *Handbooks in OR & MS*, volume 4, pages 133–173. North Holland Amsterdam, 1993.

[84] A. Federgruen and Y.-S. Zheng. An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems. *Operations Research*, 40:808–813, 1992.

[85] A. Federgruen and P. Zipkin. An efficient algorithm for computing optimal (s, S) policies. *Operations Research*, 32:1268–1285, 1984.

[86] A. Federgruen and P.H. Zipkin. Approximations of dynamic multilocation production and inventory problems. *Management Science*, 30:69–84, 1984.

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

[88] G.J. Feeney and C.C. Sherbrooke. The inventory policy under compound Poisson demand. *Management Science*, 12:391–411, 1966.

[89] A. Fink and T. Reiners. Modeling and solving the short-term car rental logistics problem. *Transportation Research*, Part E 42:272–292, 2006.

[90] P. Gaiardelli, N. Saccani, and L. Songini. Performance measurement of the after-sales service network - evidence from the automotive industry. *Computer in Industry*, 58:698–708, 2007.

[91] V. Gaur and D. Honhon. Assortment planning and inventory decisions under a locational choice model. *Management Science*, 52:1528–1543, 2006.

[92] D.P. Jr. Gaver. On base-stock level inventory control. *Operations Research*, 7:689–703, 1959.

[93] S. Gavirneni. Periodic review inventory control with fluctuating purchasing costs. *Operations Research Letters*, 32:374–379, 2004.

[94] Y. Gerchak, R.K. Cho, and S. Ray. Coordination of quantity and shelf-retention timing in the video movie rental industry. *IIE Transactions*, 38:525–536, 2006.

[95] P. Glasserman and Y. Wang. Leadtime-inventory trade offs in assemble-to-order systems. *Operations Research*, 46:858–871, 1998.

[96] S.K. Goyal. Joint ordering of two products with lost sales permitted. *Engineering Costs and Production Economics*, 16:209–212, 1989.

[97] J. Grahovac and A. Chakravarty. Sharing and lateral transshipment of inventory in a supply chain with expensive low-demand items. *Management Science*, 47:579–594, 2001.

[98] S.E. Grasman, T.L. Olsen, and J.R. Birge. Setting basestock levels in multi-product systems with setups and random yield. *IIE Transactions*, 40:1158–1170, 2008.

[99] S.C. Graves. A multiple-item inventory model with a job completion criterion. *Management Science*, 28:1334–1337, 1982.

[100] W. Groeneveld. Collectiebeleid en collectievorming - Een slinger van Foucault? Technical report, Probiblio, Hoofddorp, 2005.

[101] D. Gupta. The (Q, r) inventory system with an unreliable supplier. *INFOR*, 34:59–76, 1996.

[102] C.Z. Gurgur and T. Altiok. Analysis of decentralized multi-product pull systems with lost sales. *Naval Research Logistics*, 54:357–370, 2007.

[103] A.Y. Ha. Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, 43:1093–1103, 1997.

[104] A.Y. Ha. Stock rationing in an $M/E_k/1$ make-to-stock queue. *Management Science*, 46:77–87, 2000.

[105] G. Hadley and T.M. Whitin. *Analysis of inventory systems.* Prentice-Hall, Englewood Cliffs, New York, 1963.

[106] G. Harpaz, W.Y. Lee, and R.L. Winkler. Optimal output decisions of a competitive firm. *Management Science*, 28:589–602, 1982.

[107] F. Harris. How many parts to make at once. *Factory, The Magazine of Management*, 10:135–136, 1913.

[108] W.H. Hausman. On optimal repair kits under a job completion criterion. *Management Science*, 28:1350–1351, 1982.

[109] W.H. Hausman, H.L. Lee, and A.X. Zhang. Joint demand fulfillment probability in a multi-item inventory system with independent order-up-to policies. *European Journal of Operational Research*, 109:646–659, 1998.

[110] A.C. Hax and D. Candea. *Production and Inventory Management.* Prentice-Hall, Englewood Cliffs, 1984.

[111] D. Heeremans and L.F. Gelders. Multiple period repair kit problem with a job completion criterion: A case study. *European Journal of Operational Research*, 81:239–248, 1995.

[112] A. Henderson. The library collection failure quotient: the ratio of interlibrary borrowing to collection size. *The Journal of Academic Librarianship*, 26:159–170, 2000.

[113] R.M. Hill. Numerical analysis of a continuous-review lost-sales inventory model where two orders may be outstanding. *European Journal of Operational Research*, 62:11–26, 1992.

[114] R.M. Hill. Continuous review lost sales inventory models where two orders may be outstanding. *International Journal of Production Economics*, 35:313–319, 1994.

[115] R.M. Hill. Order splitting in continuous reivew (Q, r) inventory models. *European Journal of Operational Research*, 95, 1996.

[116] R.M. Hill. On the suboptimality of $(S - 1, S)$ lost sales inventory policies. *International Journal of Production Economics*, 59:387–393, 1999.

[117] R.M. Hill. Continuous-review, lost-sales inventory models with Poisson demand, a fixed lead time and no fixed order cost. *European Journal of Operational Research*, 176:956–963, 2007.

[118] R.M. Hill and S.G. Johansen. Optimal and near-optimal policies for lost sales inventory models with at most one replenishment order outstanding. *European Journal of Operational Research*, 169:111–132, 2006.

[119] A. Hsu and Y. Bassok. Random yield and random demand in a production system with downward substitution. *Operations Research*, 47:277–290, 1999.

[120] C.-F. Hsueh and M.-S. Chang. Equilibrium analysis and corporate social responsibility for supply chain integration. *European Journal of Operational Research*, 190:116–129, 2008.

[121] J. Hu, E. Watson, and H. Schneider. Approximate solutions for multi-location inventory systems with transshipments. *International Journal of Production Economics*, 97:31–43, 2005.

[122] W.-T. Hu, S.-L. Kim, and A. Banerjee. An inventory model with partial backordering and unit backorder cost linearly increasing with the waiting time. *European Journal of Operational Research*, 197:581–587, 2008.

[123] W.T. Huh, G. Janakirman, J.A. Muckstadt, and P. Rusmevichientong. An adaptive algorithm for finding the optimal base-stock policy in lost sales inventory systems with censored demand. *Mathematics of Operations Research*, 34:397–416, 2009.

[124] W.T. Huh, G. Janakirman, J.A. Muckstadt, and P. Rusmevichientong. Asymptotic optimality of order-up-to policies in lost sales inventory systems. *Management Science*, 55:404–420, 2009.

[125] E. Ignall and A.F. Veinott Jr. Optimality of myopic inventory policies for several substitute products. *Management Science*, 15:284–304, 1969.

[126] S.M.R. Iravani and K.L. Luangkesorn. A general decomposition algorithm for parallel queues with correlated arrivals. *Queueing Systems*, 47, 2004.

[127] K.P.S. Isotupa. An (s, Q) Markovian inventory system with lost sales and two demand classes. *Mathematical and Computer Modelling*, 43:687–694, 2006.

[128] A.A. Jagers and E.A. Van Doorn. On the continued Erlang loss function. *Operations Research Letters*, 5:43–46, 1986.

[129] G. Janakiraman and J.A. Muckstadt. *Periodic review inventory control with lost sales and fractional lead times.* PhD thesis, Stern School of Business, New York University, 2004.

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

[131] G. Janarikaman, S. Seshadri, and J.G. Shanthikumar. A comparison of the optimal costs of two canonical inventory systems. *Operations Research*, 55:866–875, 2007.

[132] S.G. Johansen. Pure and modified base-stock policies for the lost sales inventory system with negligible set-up costs and constant lead times. *International Journal of Production Economics*, 71:391–399, 2001.

[133] S.G. Johansen. Base-stock policies for the lost sales inventory system with Poisson demand and Erlangian lead times. *International Journal of Production Economics*, 93–94:429–437, 2005.

[134] S.G. Johansen and R.M. Hill. The (r, Q) control of a periodic-review inventory system with continuous demand and lost sales. *International Journal of Production Economics*, 68:279–286, 2000.

[135] S.G. Johansen and A. Thorstenson. Optimal and approximate (Q, r) inventory policies with lost sales and gamma-distributed lead time. *International Journal of Production Economics*, 30–31:179–194, 1993.

[136] S.G. Johansen and A. Thorstenson. Optimal (r, Q) inventory policies with Poisson demands and lost sales: discounted and undiscounted cases. *International Journal of Production Economics*, 46–47:359–371, 1996.

[137] S.G. Johansen and A. Thorstenson. The (r, q) policy for the lost-sales inventory system when more than one order may be outstanding. Technical Report L-2004-03, Aarhus School of Business, 2004.

[138] S.G. Johansen and A. Thorstenson. Pure and restricted base-stock policies for the lost-sales inventory system with periodic review and constant lead times. Technical report, Aarhus University, 2008.

[139] E.L. Johnson. On (s, S) policies. *Management Science*, 18:80–101, 1968.

[140] D.H. Johnston. Primary supplier relationships for all inventory items can reduce a hospital's operating costs. *Hospital Materials Management*, 17:17–18, 1992.

[141] F.R. Johnston, J.E. Boylan, and E.A. Shale. An examination of the size of orders from customers, their characterisation and the implications for inventory control of slow moving items. *Journal of the Operational Research Society*, 54:833–837, 2003.

[142] M.R.A. Jokar and M. Seifbarghy. Cost evaluation of a two-echelon inventory system with lost sales and approximately normal demand. *Scientia Iranica*, 13:105–112, 2006.

[143] B.-R. Jung, B.-G. Sun, J.-S. Kim, and S.-E. Ahn. Modeling lateral transshipments in multiechelon repairable-item inventory systems with finite repair channels. *Computers & Operations Research*, 30:1401–1417, 2003.

[144] S. Kalpakam and G. Arivarignan. A continuous review inventory system with semi-Markovian demands. *Mathematical and Computer Modelling*, 11:1172–1177, 1988.

[145] S. Kalpakam and G. Arivarignan. A lost sales inventory system in a random environment. *Stochastic Analysis and Applications*, 7:367–385, 1989.

[146] S. Kalpakam and G. Arivarignan. (s, S) inventory systems with lost sales and Markov renewal demands. *Mathematical and Computer Modeling*, 12:1511–1520, 1989.

[147] S. Kalpakam and G. Arivarignan. A lost sales inventory system with multiple reorder levels. In R. Subramanian M.J. Beckmann, M.N. Gopalan, editor, *Stochastic Processes and Their Applications*, pages 241–252. Berlin, Springer, 1991.

[148] S. Kalpakam and K.P. Sapna. A control policy of an inventory system with compound Poisson demand. *Stochastic Analysis and Applications*, 11:459–482, 1993.

[149] S. Kalpakam and K.P. Sapna. A modified (s, S) inventory system with lost sales and emergency orders. *Opsearch*, 30:321–335, 1993.

[150] S. Kalpakam and K.P. Sapna. A two reorder level inventory system with renewal demands. *European Journal of Operational Research*, 84:402–415, 1995.

[151] S. Kalpakam and K.P. Sapna. A lost sales $(S-1, S)$ perishable inventory system with renewal demand. *Naval Research Logistics*, 43:129–142, 1996.

[152] S. Kalpakam and K.P. Sapna. A lost sales inventory system with supply uncertainty. *Computers and Mathematics with Applications*, 33:81–93, 1997.

[153] P.B. Kantor. *Objective performance measures for academic and research libraries.* Association of Research Libraries, Washington, D.C., 1984.

[154] B.A. Kapalka, K. Katircioglu, and M.L. Puterman. Retail inventory control with lost sales service constraints, and fractional lead times. *Production and Operations Management*, 8:393–408, 1999.

[155] S. Karlin and H. Scarf. Inventory models of the Arrow-Harris-Marschak type with time lag. In K. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the Mathematical Theory of Inventory and Production*, chapter 10. Stanford University, Stanford, Ca, 1958.

[156] U.S. Karmarkar. The multilocation multiperiod inventory problem: bounds and approximations. *Management Science*, 33:86–94, 1987.

[157] W. Karush. A queueing model for an inventory problem. *Operations Research*, 5:693–703, 1957.

[158] J.-P. Kenné, A. Gharbi, and M. Beit. Age-dependent production planning and maintenance strategies in unreliable manufacturing systems with lost sale. *European Journal of Operational Research*, 278:408–420, 2007.

[159] W.J. Kennedy, J.W. Patterson, and L.D. Fredendall. An overview of recent literature on spare-parts inventories. *International Journal of Production Economics*, 76:201–215, 2002.

[160] E. Kim and M.P. Van Oyen. Dynamic scheduling to minimize lost sales subject to set-up costs. *Queueing Systems*, 29:193–229, 1998.

[161] J.S. Kim, S. Hur, and T.Y. Kim. A multiechelon repairable item inventory system with lateral transshipment and a general repair time distribution. *Informatica*, 17:381–392, 2006.

[162] P. Köchel. On queueing models for some multi-location problems. *International Journal of Production Economics*, 45:429–433, 1996.

[163] P Köchel. A survey on multi-location inventory models with lateral transshipments. In S. Papachristos and I. Ganas, editors, *Inventory modelling in production and supply chains*, pages 183–207, University of Ioannina, Greece, 1998. Proceedings of the Third ISIR Summer School.

[164] A.G. Kök and M.L. Fisher. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55:1001–1021, 2007.

[165] G. Koole and J. Talim. Exponential approximation of multi-skill call centers architecture. In *Proceedings of QNETs 2000*, volume 23, pages 1–10, 2000.

[166] A.A. Kranenburg and G.J. van Houtum. Cost optimization in the $(S-1, S)$ lost sales inventory model with multiple demand classes. *Operations Research Letters*, 35:493–502, 2007.

[167] G.N. Krieg and H. Kuhn. A decomposition method for multi-product kanban systems with setup times and lost sales. *IIE Transactions*, 34:613–625, 2002.

[168] G.N. Krieg and H. Kuhn. Analysis of multi-product kanban systems with state-dependent setups and lost sales. *Annals of Operations Research*, 125:141–166, 2004.

[169] A. Krishnamoorthy and M.E. Islam. (s, S) inventory system with postponed demands. *Stochastic Analysis and Applications*, 22:827–842, 2005.

[170] A. Kukreja and C.P. Schmidt. A model for lumpy demand parts in a multi-location inventory system with transshipments. *Computers & Operations Research*, 32:2059–2075, 2005.

[171] A. Kukreja, C.P. Schmidt, and D.M. Miller. Stocking decisions for low-usages items in a multilocation inventory system. *Management Science*, 47:1371–1383, 2001.

[172] E. Kutanoglu. Insights into inventory sharing in service parts logistics systems with time-based service levels. *Computers & Industrial Engineering*, 54:341–358, 2008.

[173] S.D. Lapierre and A.B. Ruiz. Scheduling logistic activities to improve hospital supply systems. *Computers & Operations Research*, 34:624–641, 2007.

[174] M.A. Lariviere and E.L. Porteus. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science*, 45:346–363, 1999.

[175] H.-S. Lau and A.H.-L. Lau. Estimating the demand distributions of single-period items having frequent stockouts. *European Journal of Operational Research*, 92:254–265, 1996.

[176] H.L. Lee. A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, 33:1302–1316, 1987.

[177] J.E. Lee and Y. Hong. A stock rationing policy in a (s, S)-controlled stochastic production system with 2-phase Coxian processing times and lost sales. *International Journal of Production Economics*, 83:299–307, 2003.

[178] W.-C. Lee, J.-W. Wu, and J.-W. Hsu. Computational algorithm for inventory model with a service level constraint, lead time demand with the mixture of distributions and controllable negative exponential backorder rate. *Applied Mathematics and Computation*, 175:1125–1138, 2006.

[179] Y.H. Lee, J.W. Jung, and Y.S. Jeon. An effective lateral transshipment policy to improve service level in the supply chain. *International Journal of Production Economics*, 106:115–126, 2007.

[180] K.-N.F. Leung. Technical note: A use of the complete squares method to solve and analyze a quadratic objective function with two decision variables exemplified via a deterministic inventory model with a mixture of backorders and lost sales. *International Journal of Production Economics*, 113:275–281, 2008.

[181] R. Levi, G. Janakiraman, and M. Nagarajan. A 2-approximation algorithm for stochastic inventory control models with lost sales. *Mathematics of Operations Research*, 33:351–374, 2008.

[182] Q. Li and A.Y. Ha. Reactive capacity and inventory competition under demand substitution. *IIE Transactions*, 40:707–717, 2008.

[183] S.-K. Liang, P. Chu, and K.-L. Yang. Improved periodic review inventory model involving leadtime with crashing components and service level. *International Journal of Systems Science*, 39:421–426, 2008.

[184] J. Little and B. Coughlan. Optimal inventory policy within hospital space constraints. *Health Care Management Science*, 11:177–183, 2008.

[185] J. Liu and C.-G. Lee. Evaluation of inventory policies with unidirectional substitutions. *European Journal of Operational Research*, 182:145–163, 2007.

[186] E.J.Jr Lodree. Advanced supply chain planning with mixtures of backorders, lost sales, and lost contract. *European Journal of Operational Research*, 181:168–183, 2007.

[187] Y. Lu. Performance analysis for assemble-to-order systems with general renewal arrivals and random batch demands. *European Journal of Operational Research*, 185:635–647, 2008.

[188] Y. Lu, J.-S. Song, and D.D. Yao. Order fill rate, leadtime variability, and advance demand information in an assemble-to-order system. *Operations Research*, 51:292–308, 2003.

[189] Y. Lu, J.-S. Song, and D.D. Yao. Backorder minimization in multiproduct assemble-to-order systems. *IIE Transactions*, 37, 2005.

[190] B. Maddah and E.K. Bish. Joint pricing, assortment, and inventory decisions for a retailer's product line. *Naval Research Logistics*, 54:315–330, 2007.

[191] S. Mahajan and G. Van Ryzin. Inventory competition under dynamic consumer choice. *Operations Research*, 49:646–657, 2001.

[192] S. Mahajan and G. Van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49:334–351, 2001.

[193] S. Mahajan and G. Van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49:334–351, 2001.

[194] J.W. Mamer and A. Shogan. A constrained capital budgeting problem with application to repair kit selection. *Management Science*, 33:800–806, 1987.

[195] J.W. Mamer and S.A. Smith. Optimizing field repair kits based on job completion rate. *Management Science*, 28:1328–1333, 1982.

[196] J.W. Mamer and S.A. Smith. Job completion based inventory systems: Optimal policies for repair kits and spare machines. *Management Science*, 31:703–718, 1985.

[197] S.T. March and G.D. Scudder. On 'optimizing field repair kits based on job completion rate'. *Management Science*, 30:1025–1028, 1984.

[198] A. McGillivrary and E.A. Silver. Some concepts for inventory control under substitutable demands. *INFOR*, 16:47–63, 1978.

[199] P. Melchiors, R. Dekker, and M.J. Kleijn. Inventory rationing in an (s, Q) inventory model with lost sales and two demand classes. *Journal of the Operational Research Society*, 51:111–122, 2000.

[200] S. Minner, E.A. Silver, and D.J. Robb. An improved heuristic for deciding on emergency transshipments. *European Journal of Operational Research*, 148:384–400, 2003.

[201] E. Mohebbi. Supply interruptions in a lost-sales inventory system with random lead time. *Computers & Operations Research*, 30:411–426, 2003.

[202] E. Mohebbi and M.J.M. Posner. A continuous-review inventory system with lost sales and variable lead time. *Naval Research Logistics*, 45:259–278, 1998.

[203] E. Mohebbi and M.J.M. Posner. Sole versus dual sourcing in a continuous-review inventory system with lost sales. *Computers Industrial Engineering*, 34:321–336, 1998.

[204] E. Mohebbi and M.J.M. Posner. A lost-sales continuous-review inventory system with emergency ordering. *International Journal of Production Economics*, 58:93–112, 1999.

[205] E. Mohebbi and M.J.M. Posner. Multiple replenishment orders in a continuous-review inventory system with lost sales. *Operations Research Letters*, 30:117–129, 2002.

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

[207] D.C. Montgomery, M.S. Bazaraa, and A.K. Keswani. Inventory models with a mixture of backorders and lost sales. *Naval Research Logistics Quarterly*, 20:255–264, 1973.

[208] S. Moon. Taking cost off supply shelf. *Modern Healthcare*, 34:26–28, 2004.

[209] P.M. Morse. *Queues, Inventories and Maintenance.* Wiley, New York, 1958.

[210] P.M. Morse. Solutions of a class of discrete-time inventory problems. *Operations Research*, 7:67–78, 1959.

[211] T.E. Morton. Bounds on the solution of the lagged optimal inventory equation with no demand backlogging and proportional costs. *SIAM Review*, 11:572–596, 1969.

[212] T.E. Morton. The near myopic nature of the lagged proportional cost inventory problem with lost sales. *Operations Research*, 19:1708–1716, 1971.

[213] J.A. Muckstadt. *Analysis and algorithms for service parts supply chains.* Springer Science and Business Media, cop, New York, 2005.

[214] M. Nagarajan and S. Rajagopalan. Inventory models for substitutable products: optimal policies and heuristics. *Management Science*, 54:1453–1466, 2008.

[215] A. Nagurney. *Supply Chain Network Economics: Dynamics of Prices, Flows, and Profit.* Edward Elgar Publishing, 2006.

[216] S. Nahmias. Simple approximations for a variety of dynamic leadtime lost-sales inventory models. *Operations Research*, 27:904–924, 1979.

[217] S. Nahmias. Demand estimation in lost sales inventory systems. *Naval Research Logistics*, 41:739–757, 1994.

[218] S. Nahmias and S.A. Smith. Optimizing inventory levels in a two-echelon retailer system with partial lost sales. *Management Science*, 40:582–596, 1994.

[219] V. G. Narayanan and A. Raman. Contracting for inventory in a distribution channel with stochastic demand and substitute products. *Production and Operations Management*, 13:46–62, 2004.

[220] J. Nathan and J. Trinkaus. Improving health care means spending more time with patients and less time with inventory. *Hospital Material Management Quarterly*, 18:66–68, 1996.

[221] P.M. Needham and P.T. Evers. The influence of individual cost factors on the use of emergency transshipments. *Transportation Research E*, 34:149–160, 1998.

[222] S. Netessine and N. Rudi. Centralized and competitive inventory models with demand substitution. *Operations Research*, 51:329–335, 2003.

[223] S. Netessine, N. Rudi, and Y. Wang. Inventory competition and incentives to back-order. *IIE Transactions*, 38:883–902, 2006.

[224] L. Nicholson, A.J. Vakharia, and S.S. Erenguc. Outsourcing inventory management decisions in healthcare: Models and application. *European Journal of Operational Research*, 154:271–290, 2004.

[225] T.E. Nisonger. A review and analysis of library availability studies. *Library Resources & Technical Services*, 51:30–49, 2007.

[226] R.D. Nobel and M. Van der Heeden. A lost-sales production/inventory model with two discrete production modes. *Stochastic Models*, 16:453–478, 2000.

[227] L.-Y. Ouyang, B.-R. Chuang, and Y.-J. Lin. Periodic review inventory models with controllable lead time and lost sales rate reduction. *Journal of the Chinese Institute of Industrial Engineers*, 22:355–368, 2005.

[228] L.-Y. Ouyang, B.-R. Chuang, and Y.-J. Lin. Effective investment to reduce lost-sales rate in a periodic review inventory model. *OR Spectrum*, 29:681–697, 2007.

[229] D. Özdemir, E. Yücesan, and Y.T. Here. Multi-location transshipment problem with capacited transportation. *European Journal of Operational Research*, 175:602–621, 2006.

[230] J.E. Pachon, E. Iakovou, C. Ip, and R. Aboudi. A synthesis of tactical fleet planning models for the car rental industry. *IIE Transactions*, 35:907–916, 2003.

[231] M. Parlar and D. Berkin. Future supply uncertainty in EOQ models. *Naval Research Logistics*, 38:107–121, 1991.

[232] M. Parlar and S. Goyal. Optimal ordering decisions for two substitutable products with stochastic demands. *OPSEARCH*, 21:1–15, 1984.

[233] B. Pasternack and Z. Drezner. Optimal inventory policies for substitutable commodities with stochastic demand. *Naval Research Logistics*, 38:221–240, 1991.

[234] D.W. Pentico and M.J. Drake. The deterministic EPQ with partial backordering: A new approach. *European Journal of Operational Research*, 194:102–113, 2009.

[235] D.W. Pentico, M.J. Drake, and C. Toews. The deterministic EPQ with partial backordering: A new approach. *Omega*, 37:624–636, 2009.

[236] E.L. Porteus. Numerical comparisons of inventory policies for periodic review systems. *Operations Research*, 33:134–152, 1985.

[237] M.J.M. Posner. Class of inventory models with customer impatience. *Naval Research Logistisc Quarterly*, 19:483–492, 1981.

[238] I. Pressman. An order-level-scheduling-period system with lost sales. *Management Science*, 17:1328–1335, 1977.

[239] K. Rajaram. Assortment planning in fashion retailing: methodology, application and analysis. *European Journal of Operational Research*, 129:186–208, 2001.

[240] K. Rajaram and C.S. Tang. The impact of product substitution on retail merchandising. *European Journal of Operational Research*, 135:582–601, 2001.

[241] U.S. Rao, J. M. Swaminathan, and J. Zhang. Multi-product inventory planning with downward substitution, stochastic demand and setup costs. *IIE Transactions*, 36:59–71, 2004.

[242] N. Ravichandran. Note on (s, S) inventory policy. *IIE Transactions*, 16:387–390, 1984.

[243] M Reiman. A new and simple policy for the continuous review lost sales inventory model. Technical Report Working paper, Bell Labs, Murray Hill, NJ, 2004.

[244] A. Reisman and X. Xu. Operation research in libraries: A review of 25 years of activity. *Operations Research*, 42:34–40, 1994.

[245] P.M. Reyes and L.M. Meade. Improving reverse supply chain operational performance: A transshipment application study for not-for-profit organizations. *The Journal of Supply Chain Management*, 42:38–48, 2006.

[246] J. Riordan. *Stochastic service systems*. Wiley, 1961.

[247] H. Rivard-Royer, S. Landry, and M. Beaulieu. Hybrid stockless: A case study; lessons for health-care supply chain integration. *International Journal of Operations & Production Management*, 22:412–424, 2002.

[248] L. W. Robinson. Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. *Operations Research*, 38:278–295, 1990.

[249] D Rosenberg. A new analysis of a lot-size model with partial back-logging. *Naval Research Logistics Quarterly*, 26:349–353, 1979.

[250] K. Rosling. The (r, Q) inventory model with lost sales. Technical report, Department of Industrial Engineering, Lund University, 1998.

[251] K. Rosling. Inventory cost rate functions with nonlinear shortage costs. *Operations Research*, 5:1007–1017, 2002.

[252] N. Saccania, P. Johansson, and M. Peronam. Configuring the after-sales service supply chain: A multiple case study. *International Journal of Production Economics*, 110:52–69, 2007.

[253] L.A. San José, J. Sicilia, and J. García-Laguna. The lot size-reorder level inventory system with customers impatience functions. *Computers & Industrial Engineering*, 49:349–362, 2005.

[254] L.A. San José, J. Sicilia, and J. García-Laguna. Analysis of an inventory system with exponential partial backordering. *International Journal of Production Economics*, 100:76–86, 2006.

[255] L.A. San-José, J. Sicilia, and J. García-Laguna. A general model for EOQ inventory systems with partial backlogging and linear shortage costs. *International Journal of Systems Science*, 40:59–71, 2009.

[256] A.E. Sarhan and B.G. Greenberg. Estimation of location and scale parameters by order statistics from singly and doubly censored samples. part 1. the normal distribution up to samples of size 10. *Annals of Mathematical Statistics*, 27:427–445, 1956.

[257] A.E. Sarhan and B.G. Greenberg. Estimation of location and scale parameters by order statistics from singly and doubly censored samples. part 2. tables for the normal distribution for samples of size $11 \leq n \leq 15$. *Annals of Mathematical Statistics*, 28:79–105, 1957.

[258] H. Scarf. The optimality of (s, S) policies in the dynamic inventory problem. In K.J. Arrow, S. Karlin, and P. Suppes, editors, *Mathematical Methods in the Social Sciences*, chapter 13. Stanford, Calif.: Stanford University Press., 1960.

[259] H.E. Scarf. Bayes solution of the statistical inventory problem. *Annals of the Institute of Statistical Mathematics*, 30:490–508, 1959.

[260] M.K. Schaefer. A multi-item maintenance center inventory model for low-demand reparable items. *Management Science*, 29, 1983.

[261] C.P. Schmidt and S. Nahmias. $(S - 1, S)$ policies for perishable inventory. *Management Science*, 31:719–728, 1985.

[262] M. Seifbarghy and M.R.A. Jokar. Cost evaluation of a two-echelon inventory system with lost sales and approximately normal demand. *International Journal of Production Economics*, 102:244–254, 2006.

[263] B. Sezen. Changes in performance under various lengths of review periods in a periodic review inventory control system with lost sales. *International Journal of Physical Distribution & Logistics Management*, 36:360–373, 2006.

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

[265] S.E. Shreve. Abbreviated proof [in the lost sales case]. In D.P. Bertsekas, editor, *Dynamic Programming and Stochastic Control*, pages 105–106. Academic Press, New York, 1976.

[266] E.A. Silver. Inventory management: A tutorial, Canadian publications, practical applications, and suggestions for future research. *Information Systems and Operational Research*, 46:15–28, 2008.

[267] E.A. Silver and H.C. Meal. A heuristic for selecting lot size quantities for the case of a deterministic time-varying demand rate and discrete opportunities

for replenishments. *Production and Inventory Management Journal*, 2nd Quarter:64–74, 1973.

[268] E.A. Silver, D.F. Pyke, and R. Peterson. *Inventory management and production planning and scheduling.* John Wiley & Sons, New York, third edition edition, 1998.

[269] E.A. Silver and S.A. Smith. A graphical aid for determining optimal inventories in a unit replenishment inventory system. *Management Science*, 24:358–359, 1977.

[270] B.K. Sinha and R.C. Clelland. Modeling for the management of library collections. *Management Science*, 22:547–557, 1976.

[271] B. Sivakumar and G. Arivarignan. A modified lost sales inventory system with two types of customers. *Quality Technology & Quantitative Management*, 5:339–349, 2008.

[272] S.A. Smith. Optimal inventories for an $(S-1,S)$ system with no backorders. *Management Science*, 23:522–528, 1977.

[273] S.A. Smith and N. Agrawal. Management of multi-item retail inventory systems with demand substitution. *Operations Research*, 48:50–64, 2000.

[274] S.A. Smith, J.C. Chambers, and E. Shlifer. Optimal inventories based on job completion rate for repairs requiring multiple items. *Management Science*, 26:849–854, 1980.

[275] J.-S. Song. On the order fill rate in a multi-item, base-stock inventory system. *Operations Research*, 46:831–845, 1998.

[276] J.-S Song. A note on assemble-to-order systems with batch ordering. *Management Science*, 46:739–743, 2000.

[277] J.-S. Song. Order-based backorders and their implications in multi-item inventory systems. *Management Science*, 48, 2002.

[278] J.-S. Song, S. Xu, and B. Liu. Order-fulfillment performance measures in an assemble-to-order system with stochastic leadtimes. *Operations Research*, 47:131–149, 1999.

[279] J.-S. Song and P.H. Zipkin. Supply chain operations: Assemble-to-order systems. In A.G. de Kok and S. Graves, editors, *Handbooks in Operations Research and Management Science*. North-Holland, 2003.

[280] S. Stidham. A last word on $L = \lambda W$. *Operations Research*, 22:417–421, 1974.

[281] G. Tagaras. Pooling in multi-location periodic inventory distribution systems. *Omega International Journal of Management Science*, 27:39–59, 1999.

[282] G. Tagaras and M.A. Cohen. Pooling in two-location inventory systems with non-negligible replenishment lead times. *Management Science*, 38:1067–1083, 1992.

[283] B. Tan and S. Karabati. Can the desired service level be achieved when the demand and lost sales are unobserved? *IIE Transactions*, 36:345–358, 2004.

[284] C.S. Tang and S Deo. Rental price and rental duration under retail competition. *European Journal of Operational Research*, 187:806–828, 2008.

[285] R.H. Teunter. The multiple-job repair kit problem. *European Journal of Operational Research*, 175:1103–1116, 2006.

[286] H. Tijms. *A first course in stochastic models*. Wiley, Chichester, 2003.

[287] H.C. Tijms and H. Groenevelt. Simple approximations for the reorder point in periodic and continuous review (s, S) inventory systems with service level constraints. *European Journal of Operational Research*, 17:175–190, 1984.

[288] A.A. Tsay and N. Agrawal. Channel conflict and coordination in the e-commerce age. *Production and Operations Management*, 13:93–110, 2004.

[289] K. Van Donselaar, T. De Kok, and Rutten W. Two replenishment strategies for the lost sales inventory model: A comparison. *International Journal of Production Economics*, 46–47:285–295, 1996.

[290] K. Van Donselaar, T. Van Woensel, R. Broekmeulen, and J. Fransoo. Inventory control of perishables in supermarkets. *International Journal of Production Economics*, 104:462–472, 2006.

[291] G.J. Van Houtum, K. Inderfurth, and W.H.M. Zijm. Materials coordination in stochastic multi-echelon systems. *European Journal of Operational Research*, 95:1–23, 1996.

[292] G.J. Van Houtum and W.H.M. Zijm. Computational procedures for stochastic multi-echelon production systems. *International Journal of Production Economics*, 23:223–237, 1991.

[293] G.J. Van Houtum and W.H.M. Zijm. On the relation between cost and service models for general inventory systems. *Statistica Neerlandica*, 54:127–147, 2000.

[294] G. Van Ryzin and S. Mahajan. On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*, 45:1496–1509, 1999.

[295] A.F. Veinott. Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes. *Operations Research*, 13:761–778, 1965.

[296] Jr. Veinott, A.F. On the optimality of (s, S) inventory policies: New conditions and a new proof. *SIAM Journal of Applied Mathematics*, 14:1067–1083, 1966.

[297] Jr. Veinott, A.F. and H.M. Wagner. Computing optimal (s, S) inventory policies. *Management Science*, 11:525–552, 1965.

[298] J.H.C.M. Verrijdt and A.G. De Kok. Distribution planning for divergent depotless two-echelon network under service constraints. *European Journal of Operational Research*, 89:341–354, 1996.

[299] V. Vincent and M. Ranton. Hospital pharmacy inventory management: Economic order quantity model with space limitation. *Hospital Material Management Quarterly*, 5:82–96, 1984.

[300] I.F.A. Vis, J. Boter, and M. Bijvank. Dynamic assignment of books to library branches. Working paper to be submitted for publication, 2009.

[301] I.M.H. Vliegen and G.J. van Houtum. Approximate evaluation of order fill rates for an inventory system of service tools. *International Journal of Production Economics*, 118:339–351, 2009.

[302] S. Wadhwa, A. Kanda, K.S. Bhoon, and Bibhushan. Impact of supply chain collaboration on customer service level and working capital. *Global Journal of Flexible Systems Management*, 7:27–35, 2006.

[303] H. Wagner and T.M. Whitin. Dynamic version of the economic lot size model. *Management Science*, 5:89–96, 1958.

[304] Q. Wang and M. Parlar. A three-person game theory model arising in stochastic inventory control theory. *European Journal of Operational Research*, 76:83–97, 1994.

[305] S.M. Ward, T. Wray, and K.E. Debus-López. Collection development based on patron requests: collaboration between interlibrary loan and acquisitions. *Library Collections, Acquisitions, & Technical Services*, 27:203–213, 2008.

[306] W.E. Wecker. Predicting demand from sales data in the presence of stockouts. *Management Science*, 24:1043–1054, 1978.

[307] J. Weinstock and D. Young. The influence of correlation of replenishment times on inventory control systems. *Operations Research*, 5:145–146, 1957.

[308] R.W. Wolff. Poisson Arrivals See Time Averages. *Operations Research*, 30:223–231, 1982.

[309] H. Wong, D. Cattrysse, and D. Van Oudheusden. Inventory pooling of repairable spare sparts with non-zero lateral transshipment time and delayed lateral transshipments. *European Journal of Operational Research*, 165:207–218, 2005.

[310] H. Wong, D. Cattrysse, and D. Van Oudheusden. Stocking decisions for repairable spare parts pooling in a mult-hub system. *International Journal of Production Economics*, 93–94:309–317, 2005.

[311] H. Wong, G.J. Van Houtum, D. Cattrysse, and D. Van Oudheusden. Simple, efficient heuristics for multi-item multi-location spare parts systems with lateral transshipments and waiting time constratints. *Journal of the Operational Research Society*, 56:1419–1430, 2005.

[312] H. Wong, G.J. Van Houtum, D. Cattrysse, and D. Van Oudheusden. Multi-item spare parts systems with lateral transshipments and waiting time constraints. *European Journal of Operational Research*, 171:1071–1093, 2006.

[313] H. Wong, D. Van Oudheusden, and D. Cattrysse. Two-echelon multi-item spare parts systems with emergency supply flexibility and waiting time constraints. *IIE Transactions*, 39:1045–1057, 2007.

[314] K. Xi, P.T. Evers, and M.C. Fu. Estimating customer service in a two-location continuous review inventory model with emergency transshipments. *European Journal of Operational Research*, 145:569–584, 2003.

[315] V.S.S. Yadavalli, N. Anbazhagan, and G. Arivarignan. A two-commodity stochastic inventory system with lost sales. *Stochastic Analysis and Applications*, 22:479–497, 2004.

[316] G.K. Yang, S.-Y. Chou, C.-Y. Hung, J.S.-J. Lin, and P. Chu. Research note on the criteria for the optimal solution of the inventory model with a mixture of partial backordering and lost sales. *Applied Mathematical Modelling*, 32:1758–1768, 2008.

[317] H. Yang and L. Schrage. Conditions that cause risk pooling to increase inventory. *European Journal of Operational Research*, 192:837–851, 2009.

[318] A. Yaspan. An inclusive solution to the inventory problem. *Operations Research*, 9:371–382, 1961.

[319] A. Yaspan. Fixed-stockout-probability order quantities with lost sales and time lag. *Operations Research*, 20:903–904, 1972.

[320] E. Yücel, F. Karaesmen, F.S. Salman, and M. Türkay. Optimizing product assortment under customer-driven demand substitution. *European Journal of Operational Research*, 199:759–768, 2009.

[321] E. Zabel. A note on the optimality of (s, S) policies in inventory theory. *Management Science*, 9:123–125, 1962.

[322] E. Zabel. Equilibrium and adjustments in noncompetitive markets: lost sales versus backlogging processes. *Engineering Cost and Product Economics*, 19:91–99, 1990.

[323] A.X. Zhang. Demand fulfillment rates in an assemble-to-order system with multiple products and dependent demands. *Production and Operations Management*, 6:309–324, 1997.

[324] R.Q. Zhang. Expected time delay in multi-item inventory systems with correlated demands. *Naval Research Logistics*, 46:671–688, 1999.

[325] X. Zhao, B. Huo, B.B. Flynn, and J.H.Y. Yeung. The impact of power and relationship commitment on the integration between manufacturers and customers in a supply chain. *Journal of Operations Management*, 26:368–388, 2008.

[326] Y.S. Zheng and A. Federgruen. Finding optimal (s, S) policies is about as simple as evaluating a single policy. *Operations Research*, 39:654–665, 1991.

[327] P.H. Zipkin. *Foundations of inventory management*. McGraw-Hill, New York, 2000.

[328] P.H. Zipkin. Old and new methods for lost-sales inventory systems. *Operations Research*, 56:1256–1263, 2008.

[329] P.H. Zipkin. On the structure of lost-sales inventory models. *Operations Research*, 56:937–944, 2008.

Samenvatting

Klantentevredenheid is tegenwoordig een belangrijk aspect voor bedrijven om mee te adverteren aangezien het lastig is om een goede positie op de markt te krijgen met de huidige concurrentiestrijd. Daarnaast heeft de komst van het internet als distributiekanaal ervoor gezorgd dat klanten steeds vaker producten vergelijken en minder loyaal zijn aan een merk of winkel. Het is daarom erg belangrijk voor bedrijven om klanten aan zich te binden. Als gevolg hiervan is er een trend waarneembaar waarin bedrijven zich minder richten op productie en kosten, maar zich met toenemende mate meer toeleggen op het service aspect. Hiervoor is een goede voorspelling van de klantenbehoefte noodzakelijk. Daarnaast moeten bedrijven flexibel kunnen inspringen op veranderingen in het vraagpatroon van klanten. Voorraden worden juist aangehouden om deze flexibiliteit op te bouwen.

Traditioneel wordt voorraadbeheer aangestuurd op kostenbeheersing in plaats van een focus op de klant. Het doel van dit proefschrift is het ontwikkelen van voorraadmodellen en daaraan gerelateerde oplostechnieken, die zich met name bezig houden met service aan klanten. Dit proces noemen wij *service inventory management*. Huidige trends uit de praktijk waarmee rekening gehouden moet worden, zijn korte doorstroomtijden in de logistieke keten, efficiency en kostenreductie, service-garantie, een toenemende concurrentiestrijd en de invloed die dit heeft op het gedrag van klanten. Aangezien huidige technieken hiervoor niet voldoende toereikend zijn, zijn nieuwe benaderingen en technieken noodzakelijk om de voorraden accuraat te beheren. In dit proefschrift worden drie verschillende settings bekeken. Iedere situatie is een deel van mijn proefschrift.

In het eerste deel staan voorraden voor *after-sales* activeiten centraal. Als een product of machine stuk gaat, dan willen klanten graag een goede service ontvangen. Dit houdt in dat ze snel en accuraat geholpen willen worden. Op het moment dat een technicus het defecte product wil repareren, moet deze over de juiste reserveonderdelen beschikken. Als dit niet het geval is, dan kan de gehele reparatie niet doorgaan. Het voorraadbeheer in deze situatie wordt gezien als een *multi-item* voorraadprobleem, aangezien de service naar de klanten afhangt van de beschikbaarheid van alle benodigde reserveonderdelen. Voorraadmodellen waarin de vraag deze *all-or-nothing* strategie kent, worden nauwelijks behandeld in de literatuur. In dit proefschrift wordt een nieuwe uitdrukking afgeleid om de service naar de klanten te berekenen. Tevens wordt er een oplossingstechniek ontwikkeld die deze uitdrukking gebruikt om ondersteuning te bieden bij beslissingen omtrent het vaststellen hoeveel onderdelen op voorraad aangehouden moeten worden. Aan

de hand van een praktijkstudie laten wij zien dat de service met 31% omhoog kan zonder meer geld te investeren in het aanhouden van extra voorraden.

In het tweede deel van dit proefschrift wordt gekeken naar *single-item* voorraadsystemen. Voor dit type systemen is het in de literatuur gebruikelijk om aan te nemen dat klanten blijven wachten op een nieuwe levering goederen op het moment dat er geen producten meer op voorraad zijn. Dit wordt *backordering* genoemd. In de praktijk blijven klanten niet altijd wachten. Sterker nog, vaak zullen klanten een ander product of een ander merk kopen, of ze kopen hetzelfde product in een andere winkel. Hoe dan ook, de oorspronkelijk vraag gaat verloren. Dit wordt *lost sales* genoemd. Voorraadsystemen met deze eigenschap zijn echter niet eenvoudig te analyseren. In dit proefschrift laten wij zien dat er een ander type bestelstrategie gekozen moet worden voor deze systemen. Het is efficienter om geen grote hoeveelheden te bestellen op het moment dat de voorraadniveaus laag zijn als gevolg van een periode met veel vraag. Zulke periodes komen namelijk niet snel achter elkaar voor. Het heeft dan ook geen zin om extra artikelen in te kopen voor de vraag die op korte termijn verloren gaat als gevolg van deze lage voorraadniveaus. In dit deel van het proefschrift is een nieuwe oplostechniek ontwikkeld om snel en efficient de bestelhoeveelheden uit te kunnen rekenen voor lost-sales voorraadsystemen. In het specifiek is gekeken naar een situatie in een ziekenhuis, waar het van cruciaal belang is dat er voldoende goederen op voorraad liggen. Aangezien voorraadbeheer niet de voornaamste prioriteit heeft binnen ziekenhuizen, bezit het voorraadsysteem van ziekenhuizen een aantal karakteristieke eigenschappen waarmee rekening gehouden moet worden. Er zijn beperkingen in de opslagcapaciteit, en daarnaast moet het systeem eenvoudig en inzichtelijk te gebruiken zijn voor de personeelsleden. Daarom hebben wij een aantal bestelregels opgesteld, die zeer eenvoudig te gebruiken zijn in de praktijk en die zorgen voor de juiste voorraadniveaus.

In het derde deel worden geen aannames gemaakt over het gedrag van klanten op het moment dat er geen voorraad meer aanwezig is. Klanten kunnen zowel blijven wachten, een substitutieproduct nemen, naar een andere winkel gaan of ze keren met lege handen naar huis. Het is een vrij complex proces om ervoor te zorgen dat al deze keuzemogelijkheden in een wiskundig model ondervangen worden. Daarom hebben wij gebruik gemaakt van simulatie om de klantentevredenheid na te bootsen. Een simulatiemodel is een prima methodiek om de presentaties van een voorraadsysteem door te rekenen. Echter voor het nemen van bevoorradingsbeslissingen zijn andere technieken nodig die sneller een afweging kunnen maken. Daarom hebben wij een vereenvoudigd model ontwikkeld waarin snel en effectief beslissingen genomen kunnen worden omtrend voorraadniveaus.

Zoals uit de drie delen blijkt, kan het erg lastig zijn om klantengedrag en klantentevredenheid te modelleren. Dit is echter wel nodig om een hoge service graad te kunnen bieden en te kunnen concurreren in de markt. In dit proefschrift hebben wij laten zien dat benaderingsmethodieken zeer efficient en effectief kunnen zijn. Het is daarbij erg belangrijk om de wensen en het gedrag van klanten centraal te stellen.

Curriculum vitae

Marco Bijvank was born in Goes, The Netherlands, on August 8, 1981. From 1993 to 1999 he attended his secondary school at the Niftarlake College in Maarssen, The Netherlands. He studied Business Mathematics and Informatics at the VU University in Amsterdam, The Netherlands, with a specialization in optimization of business processes. In 2002 he graduated cum laude on his Bachelor's thesis for NedTrain Services (part of Dutch railways). Two years later he carried out his final project at ORTEC Consulting in Gouda, The Netherlands, and obtained his Master of Science (MSc) degree cum laude on his thesis 'Shutdown Scheduling - A practical approach to handle shutdowns at refineries'. In 2004, he has been nominated for best Business Mathematics and Informatics Thesis Award.

In 2004, he started as a Ph.D. candidate at the VU University Amsterdam, where he performed research on inventory control systems. In 2008 he has been a visiting scholar for two months at the University of Aarhus in Denmark, in co-operation with Søren Johansen. During his Ph.D. trajectory, Marco Bijvank has given presentations at conferences and workshops in Europe and North America. His work has been nominated for the Philips Mathematics Prize for Ph.D. Students 2007. Next to that, he participated in several activities: teaching to students for Business mathematics and informatics and Business Administration, following courses at the LNMB (Dutch Network on Mathematics of Operations Research), organizing workshops and seminars.

Besides his activities as a Ph.D. candidate, he participated three times in the Mathematics with Industry Study Group where mathematicians collaborate to solve problems in a business environment. Furthermore, Marco Bijvank has been affiliated as employee at the Department of Communication at the VU University to organize the recruitment of new students.

From February to June 2009, Marco Bijvank has been an assistant professor at the VU University Amsterdam. He taught courses on data mining, logistics and operations research. Marco defends his thesis on October 22, 2009.