Inventory management systems

Janssen, F.B.S.L.P.

Publication date:
1998

CentER
for
Economic Research

Inventory Management Systems

control and information issues

F.B.S.L.P. Janssen

Tilburg University

Stellingen

behorende bij het proefschrift

Inventory Management Systems

control and information issues

van

F.B.S.L.P. Janssen

I

Beschouw een stationair voorraadsysteem dat bestuurd wordt door een bestelpuntmethode. De vraag wordt beschreven door een algemeen punt-proces en levertijden zijn stochastisch. Het criterium voor het bepalen van het optimale bestelpunt is het minimaliseren van de som van bestel-, voorraad-, en boetekosten, waarbij de voorraadkosten en de boetekosten recht evenredig zijn met het aantal eenheden op voorraad respectievelijk tekort per tijdseenheid. Dan wordt de noodzakelijke voorwaarde voor het optimale bestelpunt s gegeven door

$$P_3(s) = \frac{b}{b+h}$$

met $P_3(s)$ de fractie van de tijd dat de fysieke voorraad positief is, en b en h zijn respectievelijk de boete- en voorraadkosten per eenheid per tijdseenheid.
zie Hoofdstuk 2.

II

Gebruik van meer informatie over het vraagproces maakt het mogelijk om betere schattingen te geven van de performance-indicatoren van een voorraadsysteem.
zie Hoofdstuk 4.

III

De inkoopprijs heeft veel meer invloed op de leveranciersselectie dan de lengte van de levertijd.
zie Hoofdstuk 5.

IV

De P_1 servicemaat in een "order splitting" omgeving is onbruikbaar.
zie Hoofdstuk 6.

V

Delivery splitting is een win-win strategie.

zie dit proefschrift, Hoofdstuk 8, en GOLDRATT, E.M. AND J. COX (1989). *The goal.* Aldershot: Gower, page 243-244.

VI

Meer dan de helft van alle stellingen bij proefschriften is bedoeld als zijnde schertsstellingen.

VII

Stellingen zijn een ideale manier om onderzoek dat niet gerelateerd is aan de promotie te promoten

zie LI, Q., F.B.S.L.P. JANSSEN, Z. YANG, AND T.IDA (1998). ILIN: an implementation of the integer labeling algorithm for integer programming. *IEICE transaction on Fundamentals of Electronics, Communication and Computer Sciences*, **E81-A no.2**, 304-309.

VIII

Stellingen zijn vaak geen stellingen maar meningen.

IX

Stelling **IX** behorende bij het proefschrift van Jansen (1998) gaat nergens over.

zie JANSEN, J.B. (1998). *Service and inventory models subject to a delay-limit.* Proefschrift, Katholieke Universiteit Brabant, september 1998.

X

Gesteld mag worden dat stellingen over stellingen vooral geponeerd worden om een extra stelling toe te voegen aan de stellingen

XI

Stelling **X** is de laatste stelling van de stellingen behorende bij dit proefschrift die waar is.

zie HOFSTADTER, D.R. (1980). *Gödel, Escher, Bach: an eternal golden braid*, Harvard Press.

Inventory Management Systems:

control and information issues

Inventory Management Systems:

control and information issues

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Katholieke Universiteit Brabant, op gezag van de rector magnificus, prof. dr. L.F.W. de Klerk, in het openbaar te verdedigen ten overstaan van een door het college van decanen aangewezen commissie in de aula van de Universiteit op

woensdag 9 september 1998 om 16.15 uur

door

FREDERICUS BERNARDUS SIBILLA LEONARDUS PETER JANSSEN

geboren op 21 maart 1967 te Venray

PROMOTORES: Prof. dr. F.A. van der Duyn Schouten
Prof. dr. A.G. de Kok

Voor Anita and ôs mam

Voorwoord

Ik wil beginnen met het bedanken van het SOBU (SamenwerkingsOrgaan Brabantse Universiteiten) voor het subsidiëren van mijn promotieonderzoek. Vooral de nauwe betrokkenheid van Karin Leurs, Els van Loon en Marianne Wagemans stelde ik erg op prijs.

Veel dank ben ik verschuldigd aan mijn beide promotoren: Ton de Kok en Frank van der Duyn Schouten. Ton ondersteunde mij vooral met zijn briljante ideeën, en wist mij daarmee de goede richtingen op te sturen. Tijdens de afspraken met Ton overspoelde hij mij altijd met mogelijke uitbreidingen, aanknopingspunten, en valkuilen, die mij stimuleerden tot het uitwerken van de nodige verfijningen van de modellen. Verder was het Ton die mij ongeveer 6 jaar geleden tijdens het vak Industriële Wiskunde enthousiast wist te maken voor het vakgebied van de Operations Research. Frank bedank ik voor zijn gestructureerde en professionele begeleiding. Door zijn precieze manier van werken wist hij altijd de puntjes op de i te zetten, zowel mathematisch als tekstueel. Zijn inbreng heeft de kwaliteit van het proefschrift sterk verbeterd. Verder wil ik Ruud Heuts bedanken voor zijn enthousiaste begeleiding. Ruud was mijn sparringpartner wanneer ik nieuwe ideeën had. De vele uitgebreide discussies die we hadden leidden vaak tot heldere inzichten. Verder stimuleerde Ruud mij afstudeerbegeleiding op me te nemen. Zo sloeg Ruud meestal twee vliegen in één klap, namelijk hij hoefde de betreffende student zelf niet te begeleiden en hij behoedde mij er voor de voeling met te praktijk te verliezen. Daarbij mag ik natuurlijk niet vergeten te bedanken de studenten die het vertrouwen in mij hebben geschonken hen te begeleiden.

Dan wil ik nog een aantal mensen danken die hebben bijgedragen aan een leuke tijd die ik tijdens mijn AIO periode heb gehad. Daarvoor wil ik beginnen bij mijn kamergenoot Jorg Jansen. Het motto "samen uit samen thuis" was Jorg en mij (Jansen & Janssen) op het lijf geschreven. Samen congressen bijwonen, LNMB diploma halen, snooker en poolbiljart spelen en tot slot op dezelfde dag promoveren. Verder gaat er een bijzonder woord van dank uit naar Jalal Ashayeri en Leo Strijbosch met wie ik vele malen geluncht heb waar de nodige afleidende gespreksonderwerpen ter sprake kwamen. Verder was ik ook graag op de TUE bij vakgroep LBS, alhoewel dat niet aan de frequentie en regelmaat van mijn bezoeken af te lezen was. Hierbij wil ik in het bijzonder mijn kamergenote Jannet van Zante-de Fokkert bedanken.

Er zit echter ook een keerzijde aan de medaille. Er is namelijk ook een tijdstip geweest dat ik de man met de hamer tegen ben gekomen om het zo maar eens uit te drukken. Gedurende deze periode hebben drie personen een cruciale rol gespeeld in het overbruggen van deze mindere periode. Radboud Schmitz heeft me geholpen bij het vinden van de juiste prioriteiten en eisen die ik mezelf stelde ten aanzien van het proefschrift en in mijn sociale leven. Verder heeft mijn moeder een rol van onschatbare waarde gespeeld in het opnieuw opladen van energie wanneer ik tijdens deze periode weer eens in een diep dal geraakt was. Tenslotte wil ik Anita bedanken, allereerst alleen al voor er voor mij te zijn op de momenten dat ik dat nodig had, dus vooral in het laatste jaar. Maar ook daarvoor was Anita de belangrijkste factor in het schrijven van het proefschrift. Het enthousiasme waarmee ik naar mijn werk ging was grotendeels te danken aan de liefdevolle relatie die ik met Anita heb. Zonder Anita was dit proefschrift er waarschijnlijk niet geweest.

Naast deze drie personen wil ik de hele familie Swinkels en mijn zus Peggy niet ongenoemd laten in het proefschrift. Iedereen van de familie Swinkels, van Bowie tot Mam Swinkels, heeft op de een of andere manier bijgedragen aan de geborgenheid en ontspanning die ik bij "de Swinkies" vond en dus indirect aan de kwaliteit van dit proefschrift. mijn zus Peggy heeft gedurende de weg naar het proefschrift, in de dalen en toppen, de hele tijd haar medeleven getoond.

Deze laatste periode was dus een periode van bezinning, om eens goed na te denken over wat mijn eigen eisen voor dit proefschrift waren. Dit heeft dan ook geresulteerd in een van de belangrijkste ontdekkingen tijdens mijn promotie traject: "goed is goed genoeg". Nu ligt er dan ook een proefschrift waar ik zeer tevreden mee en trots op ben.

Al met al is het promotie-traject een leerproces, waar je vele malen sterker uit komt dan dat je er mee begint.

Fred Janssen

Contents

Chapter 1

Motivation and general overview

1.1 Introduction

This thesis is about the management of single product inventory systems. Inventory management involves controlling the flow of materials, supplies or finished goods into and from a stockpoint. Materials are procured from suppliers. The buffer of stock is used to satisfy customers' demand, where customers can be consumers, external companies or internal orders from other stockpoints within the company. In this thesis we focus on a local sales organisation dealing with finished goods reserved for an external market. Hence, we do not deal with specific production issues such as finite capacity problems.

Since Harris' (1913) discovery of the economic order quantity, the number of papers published on single product inventory problems is well into the thousands (Lee and Nahmias (1993)). This immediately raises the question why another thesis on single product inventory problems? Nowadays reasons exist that warrant our contribution, namely

- Sophisticated information technology makes more detailed information available.
- More powerful computers can provide fast management tools.
- There is considerable confusion about single sourcing versus multiple sourcing.
- Over the years the customer market has changed into a more erratic and diversified one, opening the way to more advanced demand management strategies.

In the following sections we will elaborate these points further and give an overview of the content of this thesis. The sequel of the introduction is an overview of Chapter 2, wherein we review some preliminary results of inventory theory.

Chapter 2 starts with a discussion on the most important characteristics of a stockpoint, which are: the demand process; the lead time process; and the control policy. Then we will define the relevant system and performance measures. In section 2.2 we will present some mathematical background, involving renewal theory, for general inventory models. In determing optimal control parameters for inventory systems mainly two directions can be distinguished: methods that minimize total relevant costs (see, e.g., Hadley and Whitin (1963), Das (1985), Johansen and Thorstenson (1993)), and methods that are based on achieving a pre-specified customer service level (see, e.g., Schneider (1981), Schneider and Ringuest (1990), de Kok (1991a, 1991b, 1991c), Tijms and Groenevelt (1984), and Tijms (1994)). Since there is some unclarity about the exact relation between those two perspectives, we will discuss this issue for the (s, Q) inventory model in section 2.3. Finally, section 2.4 discusses the fitting of mixed Erlang distributions, based on the first two moments of the associated stochastic variable. In addition useful properties are given of mixed Erlang distributed stochastic variables.

1.2 Inventory management information systems

The critical examination of the status of management information systems began three decades ago (Ackoff (1967)). A good survey on research of management information systems is given by Galliers (1992). The increasing quality of information technology makes more detailed information available for all kind of processes. Hence, it is easy to collect daily or weekly demand information instead of monthly data. This, however, has consequences for the modelling of the demand process. The probability that demand is positive during a month is often close to one, whereas the probability that demand is positive during a day may be considerably less than one. The fact that information technology makes information about demand readily available, almost for free, enables the reduction of review periods. This has the advantage that the reaction time is short in case sudden changes in demand occur, which may be very important in a dynamic world. To model such demand processes we suggest in Chapter 3 of this thesis to use a compound Bernoulli demand process.

Shortening the time unit has apparently advantages, but one could even go one step further. Instead of collecting data per time unit one could monitor each customer individually. In this way more information about the interarrival times of customers comes available, which makes it possible to model demand as a compound renewal process. Investigations in Chapter 4 show that the level of smoothness of customer interarrival patterns has a major impact on the customer service level in case the demand is modelled as a discrete time process. Apparently this way of modelling is not flexible enough to describe the demand processes in practice nowadays.

1.3 Order splitting as purchase management policy

Since the introduction of order splitting by Sculli and Wu (1981) many papers have appeared dealing with this vendor management strategy. Order splitting can be applied in combination with many inventory replenishment strategies, such as the (s, S) and (s, Q) strategy. The order splitting strategy, or multiple sourcing strategy, is the partitioning of a replenishment order among two or more suppliers.

The essence of order splitting is to reduce the variability of the supply lead times.

Most papers focus on the variance reduction of the effective lead time, where the effective lead time is defined as the minimum of all lead times. As a result of this reduction replenishment orders may be postponed, i.e. the same customer service can be achieved with a lower safety stock. In addition order splitting may decrease the average cycle stock, since the splitted replenishment orders do not arrive in one lot. Besides, these reductions in safety and cycle stock, the delivery performance and the quality and the price of the goods can be improved through competition between the vendors. On the other hand, order splitting could increase the ordering or transhipment costs. Consequently, for a fair and meaningful evaluation of order splitting we must allow extra ordering costs for the splitted replenishment orders. We now give a comprehensive literature review on order splitting.

Sculli and Wu (1981) consider two suppliers whose lead times are normally distributed with different means and variances. Via numerical integration the mean and variance of the effective lead time are computed and tabulated for various values of the lead time parameters.

Hayya, Christy and Pan (1987) motivate their research by the safety stock reductions and economic competition that can be obtained by using two vendors instead of one. They report savings of $ 4 billion in life-cycle costs of roughly $ 20 billion in total by the splitting of jet engine orders between General Electric and Pratt and Whitney in 1983. A simulation experiment demonstrates that the use of two vendors can reduce inventory investments for a given service level when demand is normally distributed and lead times are Gamma distributed.

Pan and Liao (1989) use the order splitting model to describe a Just-In-Time inventory system. They formulate an extension on the Economic Order Quantity (EOQ) model such that the replenishment orders are split equally among n deliveries with the same supplier, and provide three simple rules to determine the optimal number of deliveries. Larson's (1989) note on the article of Pan and Liao (1989) reports that when no extra ordering costs are incurred for the splitted deliveries the optimal delivery lot size is one unit.

Ramasesh (1990) also merges the JIT philosophy with the order splitting concept. Again an EOQ model is presented in which ordering costs for each partial lot are charged differently from the regular replenishment orders. Ramasesh suggest to use this order splitting model as intermediate stage in achieving the ultimate form of JIT purchasing.

Kelle and Silver (1990a, 1990b) present two papers in which they analyze order splitting among n suppliers with Weibull lead times and constant demand. Moments of the order statistics are used to give expressions for the reduction in the expected value and the variance of the total demand until the first delivery. These moments are used to give analytical expressions for the the reorder level that provides a given probability of no stockout prior to the first delivery. In the second paper also the fraction of the demand delivered directly from shelf is used as service level constraint, and expressions for the non-stockout probability prior to the first delivery are given in case the demand is stochastic. Both papers concentrate on the first delivery and give bounds within which the shortages just before the later deliveries are negligible.

Ramasesh (1991) presents theoretical concepts underlying the order splitting strategy and its implications for implementation in procurement management. It is shown that in case supply lead times are uncertain use of dual-sourcing offers savings in inventory holding and shortage costs. It is argued that when order splitting is applied only the cost of handling receipt transaction (involving receiving, incoming inspection, storage and handling of an order) tends to increase the ordering cost significantly. But the additional costs will usually be small, since they are only in terms of setting up two separate transactions, e.g., two separate shipments, two inspections. The variable costs which vary with the quantity ordered, such as freight cost per ton, inspection cost per unit, etc., will remain the same.

Ramasesh et al.(1991) analyze dual sourcing in the context of the (s, Q) inventory model with constant demand and identical stochastic lead times. Two classes of distributions for the lead times are investigated: the uniform and the exponential distributions. Closed form expressions for the sum of ordering, holding and shortage costs are given, where shortage costs are charged per unit per unit of time. In this paper Ramasesh et al. conclude that in case the uncertainty in the lead times is high and the ordering costs are low dual sourcing could be cost effective.

Zhao and Lau (1992) consider two non-identical suppliers. They note that by selecting a second supplier with suitably higher average lead time than the first, the average inventory level can be substantially reduced. Savings from cycle stock costs often exceed savings from the safety stock costs, which is in contrast with findings in earlier papers. Simulation was used to substantiate their findings.

Hong and Hayya (1992) formulate a mathematical programming problem to solve the optimal number of suppliers from the deterministic EOQ model, assuming that the aggregate ordering cost incurred by order splitting is a non-decreasing function of the number of deliveries.

Lau and Zhao (1993) present a procedure for determining the optimal ordering policy, including the optimal split of the replenishment order, for the dual sourcing strategy. They consider stochastic demand and stochastic lead times, and minimize the sum of ordering and holding costs subject to a maximum stock-out risk. Again they report that the major advantage of order splitting is the reduction in cycle stock and not in safety stock. Furthermore, they conclude that the supplier with the largest average lead time should be allocated the larger share.

The paper of Lau and Lau (1994) considers the following question: if a second supplier offers a lower price but has a poorer lead time performance than the first supplier, how should the buyer use the two suppliers? For constant demands they minimize the sum of ordering, holding and shortage costs, where shortages are charged per unit short. Numerical sensitivity analyses indicate that the co-ordination of two suppliers is beneficial in a wide variety of situations where the various inventory parameters have intermediate (i.e. neither very high nor very low) values.

Chiang and Benton (1994) present a theoretical investigation into the consequences of different cost structures on the relative performance of sole sourcing versus dual sourcing. The computational results indicate that dual sourcing provides a better fill rate service than sole sourcing. Moreover, Chiang and Benton state that dual sourcing results in larger order quantities than sole sourcing, which indicates that attractive quantity discounts may not be in jeopardy.

Guo and Ganeshan (1995) investigate how the mean and variance of the effective lead time changes with the number of suppliers. Using analytical results on order statistics they give simple decision rules for finding the optimal number of suppliers.

Chiang and Chiang (1996) examine the possibility of arrangement for multiple deliveries in each order cycle at one single supplier. It is shown that splitting the replenishment order into multiple deliveries can significantly reduce the holding costs, especially when the interarrival times between deliveries are determined optimally.

Finally Hill (1996) compares three models concerning multiple deliveries, namely: sole sourcing, multiple sourcing and sole sourcing but ordering n times as often (i.e. replenishment order of size Q/n). Besides the observation that multiple sourcing could be economically better than sole sourcing, the major conclusion is that placing n replenishment orders of size Q/n at different points in time at one single supplier will result in lower operating costs than placing n orders of Q/n at the same point in time on different identical suppliers.

In Chapter 5 of this thesis we first explore the question raised by Lau and Lau (1994): to what extent should we use a second supplier offering different price and lead time characteristics? A drawback of order splitting, in general, is that all suppliers are used simultaneously every replenishment moment. But when enough stock is available at an order epoch the fast (and expensive) suppliers are not required. In the two-supplier model presented in Chapter 5 we consider a rigid supplier, that provides quantity discounts or a low purchase price for the majority of the purchase volume, and a flexible supplier is used for the possibility to react to short term changes in demand. For example, the largest share of the purchase volume is placed at a manufacturer and the remaining share at a distributor or wholesaler. General supply agreements are made with the main supplier to deliver a fixed quantity Q every review period. It is assumed that the lead time is deterministic. At review epochs the inventory position (defined as the stock on hand plus outstanding orders minus backorders) is evaluated. When the inventory position is below the order-up-to level

S, an order is placed at the second supplier such that the inventory position is raised to the order-up-to level. The replenishment orders from the second supplier will arrive after also deterministic lead times. It turns out that only for situations in which customer demand is erratic it is profitable to purchase a large share of the purchase volume from the flexible and expensive supplier. This conclusion has considerable impact for companies which focus on flexibility, moreover it explains why purchasers are focusing on purchase cost above deliver flexibility.

In Chapter 6 we go into detail on the order splitting strategy. We will present an (s, Q) inventory model with order splitting. Demand is modelled as a compound renewal process, and lead times of the suppliers are independent and identically distributed random variables. Regarding the literature, most papers on order splitting consider constant demand models or consider at most two suppliers. In that sense these models are special cases of the model presented in Chapter 6.

Under these general assumptions we derive very good approximate expressions for the most commonly used performance measures of inventory systems (which will be defined in section 2.1.4). The service measure mostly used in an (s, Q) inventory model with order splitting is the non-stockout probability during a replenishment cycle. We will discuss the applicability of this service measure in an order splitting environment. When order splitting is applied, replenishment orders are split equally among n suppliers. We will argue that the non-stockout probability during a replenishment cycle in an order splitting environment requires a redefinition due to the structure of the replenishment cycle. Focusing on only the first arriving partial delivery can lead to undesirable performance of the inventory system.

Finally, we present an approximation algorithm for computing optimal values for the control parameters, s, Q and n, given parameters for the underlying demand and lead time process. We experienced that in practice only the first two moments of the underlying processes can be accurately estimated from the available data. The algorithm developed in Chapter 6 can be applied in many different practical settings. We concentrate on the impact of the problem parameters on the optimal number of suppliers. For this purpose it is sufficient to consider the case of identically distributed lead times.

1.4 Demand management

Since Forresters book Industrial Dynamics (1961) many papers have been published on the control of multi-stage logistics chains. The problems signalled by Forrester with respect to amplification of demand fluctuations upstream in the logistic chain have been understood widely and Material Requirements Planning (MRP) (see, e.g., Orlicky (1975)) and Distribution Requirements Planning (DRP) (see, e.g., Martin (1990)), is used throughout industry to eliminate this amplification as much as possible. Yet, a closer look reveals that these systems typically operate within industrial and retail organizations, but seldom, if ever, across different organizations in the logistics chain. Although it is claimed by various authors, such as Martin (1990), that coupling of MRP systems used by end product manufacturers with DRP systems of component manufacturers should solve or at least alleviate these problems, it is still rare that such an approach is implemented.

The study presented in Chapters 7 and 8 is motivated by the article of Janssen and de Kok (1996), in which problems are discussed arising at a European electronics manufacturer supplying the global market. One of the major problems concerning operational control of the supply chain was the so-called "big order" issue. The company was regularly confronted with unexpected big orders for particular products, while the consumer demand faced by the customers of the company was relatively stable. Apparently the stable consumer demand was changed into erratic customer demand.

This phenomenon can be explained from the fact that sales people are usually driven by "end of the month turn-over targets", while the buyers usually are price-driven (see the discussion above on the lead time flexibility versus purchase costs). Hence, the communication between the suppliers and buyers is characterised by offerings and acceptances from time to time of high discounts, leading to (much) larger transactions than those which are based on the basic paradigm of inventory control, i.e. balancing ordering costs against inventory holding costs. When we take this type of transactions for granted, it is clear that there is room on both sides (supplier and buyer) to consider splitting of deliveries. The buyer has advantage of splitting the replenishment orders because the whole order is not needed immediately and therefore the average stock can be reduced (cf. order splitting). The supplier has advantage because this can reduce (as we will show) the safety stock on the suppliers side. It is noteworthy that too high inventory levels at the retailers also will result in high inventories at the manufacturer, which in turn requires high safety stock levels. Of course the best way to break this vicious circle is to change the short term objectives of the managers involved. This, however, is not easily established since it requires far reaching organisational changes. Order splitting seems to be a good myopic concept to decrease the negative influences of this non-optimal situation. It should be noted that in practice order splitting occurs on an ad hoc basis, mostly in situations where a customer order cannot be filled immediately. We advocate the use of order splitting on a routine basis.

In Chapters 7 and 8 we address the problem from the suppliers' point of view. The suppliers' supply chain consists of a factory, a regional distribution center, and local sales organizations with local stock. We model this supply chain as a two-echelon divergent system. The main idea is to smooth demand at the local sales organizations by offering customers different service conditions for large orders and small orders. Although this might not be in the customers advantage, we claim that the savings made by the supplier are of such a magnitude that this provides funding for discounts to customers to stimulate the acceptance of a differentiated customer service policy. In order to smooth the delivery flow we suggest two demand management strategies, namely large order overflow and delivery splitting.

A lot of research has been devoted to supply chain management in the last five years (see, for example, Lee and Billington (1992, 1993)). Their approach is based on a methodology similar to our research: a combination of empirical research and the application of quantitative models. Another related paper by Vastag et al. (1994) gives a general overview of the costs involved in the management of supply chains. The literature on quantitative modelling of supply chains is huge. For an excellent overview of the research in this area we refer to Federgruen (1993), Axsäter (1993), and Diks et al. (1997). It should be noted that the research reviewed in Federgruen (1993) and Axsäter (1993) is focused on minimization of total costs, consisting of holding, ordering, and penalty costs, whereas Diks et al. (1997) focused on cost minimization subject to a service level constraint.

A subject related to our work is risk pooling as described in Eppen and Schrage (1981), and Jönsson and Silver (1987). Risk pooling arises in our situation when we re-route large customer orders to an alternative source to stabilize demand at the stock points of the sales organizations. We will not deal with the economic theory on discount policies, as discussed in Silver and Peterson (1985) and Tersine (1994). Merely we would like to draw attention to the impact of discount policies on demand variability and to give an estimate of the cost reduction caused by employing a strategy aimed at stabilizing customer demand. This cost reduction can be used to give discounts to customers that operate according to the service strategy of the supplier.

1.4.1 Large order overflow

Large order overflow will be discussed in Chapter 7 and is characterised as follows. In general, a stockpoint in a multi-echelon distribution chain satisfies all customers that arrive at that particular stockpoint, where customers are defined as the external customers as well as replenishment orders of downstream stockpoints in the distribution chain. However, in case large order overflow is applied, customers with large demand are not satisfied by the stockpoint at which they arrive, but by an upstream stockpoint. Thus, for each stockpoint a customer order threshold quantity and an alternative source are given such that customers with demand larger than that threshold quantity are satisfied by the alternative source instead of stockpoint itself.

The analysis of the large order overflow case requires the analysis of a multi-echelon system. In literature, large order overflow is also known as "transaction cut-off", see Silver(1970), or "break quantities", see Klein and Dekker (1997), and Dekker et al.(1997). Klein and Dekker discuss a tactical optimization problem that arises when orders can be delivered from any stockpoint in the distribution system. They state that using break quantities is profitable in situations where demand is erratic (i.e. occasional very large demand transactions interspersed among a majority of small transactions). Dekker et al.(1997) discuss large order overflow in a 1-warehouse, N-retailers distribution system with (R, S) policies. Under the assumption that the stock at the warehouse is reserved for large orders only, and demand during the lead times is normally distributed, they derive an expression for the inventory costs. With this expression they provide insight into the impact of large order overflow on the inventory costs. In Chapter 7 we will discuss a heuristic algorithm to determine the control policies in a divergent (s, Q) multi-echelon system with large order overflow when demand at the stockpoints is described by a compound Poisson process. A trade-off between a decrease in ordering and holding costs versus an increase in transhipment costs is used to decide to use large order overflow or not. Moreover, it is possible for the management to investigate the costs of using a certain stockpoint i by evaluating the situation that all customers arriving at stockpoint i are diverted (i.e. the stockpoint becomes redundant).

1.4.2 Delivery splitting

Delivery splitting is dealt with in Chapter 8 and is characterised as follows. Demands of a size exceeding a certain threshold are not delivered in one single batch, even in case the inventory level is sufficiently large. The customer receives only a limited quantity, equal to the threshold quantity, at a time. If the demand size is larger than the threshold quantity, starting at the demand epoch, an amount of the threshold quantity is delivered in a number of deliveries which are a fixed time unit apart. Consequently, all quantities delivered are equal to the threshold quantity except possibly the last.

Delivery splitting is closely related to order splitting. When the customers apply order splitting at one supplier, as has been proposed by Chiang and Chiang (1996), this means that the supplier faces an special form of delivery splitting. However, it should be noted that delivery splitting is not in complete correspondence with the order splitting from the buyers point of view, because in the latter the number of deliveries is determined by the inventory situation of the individual buyer, while in the former the value of the threshold value and the distance between two subsequent deliveries of one and the same order are to be set by the supplier, based on possible safety stock reductions on suppliers side.

We have argued before that the problem of erratic demand is caused by ordering policies, which can be considered to be non-rational from an inventory management point of view. Yet, these policies may be quite rational from the perspective of short term cost minimization or other incentives. Because we advocate a long-term perspective, we follow the line of thought advocated in the Just In Time literature (see, e.g., Hall (1983)), where all waste should be avoided. Apparently such non-rational policies increase the amount of stock held in the supply chains, which is counterproductive from a long-term perspective.

Chapter 2

Terminology and background for inventory models

2.1 Description of inventory models

The inventory systems in this thesis are characterised by the demand process, the replenishment process, the replenishment policy, and by the deliver policy. It is assumed that information about these characteristics is known, for instance, the first two moments or the shape and parameters of the distribution function. Hence, we do not discuss forecasting issues or try to find the optimal replenishment policy itself. The emphasis is on optimizing the parameters of a given control policy. Therefore, we need specifications of costs and quality related system variables of the inventory system. In section 2.1.1, we will discuss the structure and the modelling of the demand process. Section 2.1.2 deals with the replenishment process. Herein, we address a well-known problem about the non-crossing of consecutive orders. Next, in section 2.1.3, we will give an overview of the most common replenishment policies. Finally, in section 2.1.4, the definitions are given of system variables and performance measures which are related to the costs and quality of the inventory system.

2.1.1 Demand process

It is almost impossible to predict with certainty the time patterns of demand. Therefore it is assumed that customers place orders at the stock keeping unit according to a stochastic process. In practice, the stochastic process associated with the demand will always change with time. Most products have a traditional product life cycle that consists of four phases (see, for example, Stahl and Grigsby (1992)), namely: the introduction phase, the growing phase, the saturation phase and the declining phase. In this thesis, we will consider products for which demand characteristics slowly change with time, i.e. for the mathematical modelling we may assume a stationary stochastic inventory system.

Customer orders have no due dates. This means that a customer placing an order at the stockpoint requires immediate delivery. When a customer arrives at the stockpoint with a demand larger than the available stock on hand, the available stock on hand is delivered directly whereas the remaining demand is backlogged. Note that in multi-echelon systems this could lead to difficulties in modelling replenishment orders, since mostly it is assumed that replenishment orders arrive in one batch.

2.1.2 Lead times

The lead time is defined as the time between initiation and receipt of a replenishment order. Lead time uncertainty is known to have a critical impact on the performance and costs of an inventory system (cf. Gross and Soriano (1969)). However, analytical analysis with independent identically distributed lead times are often very cumbersome, even when lead times are exponentially distributed (see, Galliher et al.(1959)). As long as there is never more than one single order outstanding, no theoretical difficulties are encountered. Unfortunately, in most inventory model there is a positive probability that more than one order is outstanding. In the latter case it is not possible to represent lead times as independent random variables. The difficulty is that if orders are not permitted to cross, successive lead times are dependent random variables. Hadley and Whitin (1963) give a comprehensive discussion about the order crossing problem. They also state that when the intervals between successive replenishment orders is large with respect to the magnitude of the actual lead times the probability of order crossing is negligible.

The dependency of lead times is not a restriction for inventory theory. However, the validation of the analytical model with simulation becomes more cumbersome when lead times of replenishment order are not allowed to cross. Therefore, a number of mechanisms are developed to generate dependent and identically distributed lead times which do not cross.

Kaplan (1970) investigated the characterisation of optimal policies for a dynamic inventory problem when the time lag in delivery is a discrete random variable. The approach assumes a mechanism for the arrival of orders, which ensures that orders never cross, while in general lead times will be dependent. Zipkin (1986a) extended the mechanism of Kaplan (1970) for continuous time models.

Heuts and de Klein (1995) give a mechanism which can not be incorporated in the mechanism of Zipkin. They introduced the following dependency with respect to the successive lead times. If at an order epoch t all orders placed previously have arrived, then the lead time of this order is called a start-up order and is a discrete random variable. On the other hand, if at an order epoch t at least one outstanding order did not arrive yet, then this order is called a follow up order and will arrive simultaneously with the first order not arrived yet (necessarily a start-up order).

Diks and van der Heijden (1997) have developed a mechanism that generates random lead times, with pre-specified first two moments, which are dependent but do not cross. The lead time is modelled as the sum of the sojourn time in a $GI|G|1$-queue plus a deterministic pipeline time. The arrivals of customers at the queue correspond to placing a replenishment order, whereas the completion plus the deterministic pipeline time corresponds to an arrival. As a possible interpretation of this mechanism, the sojourn time of the queue can be seen as a production time, and the deterministic pipeline time as the transhipment time.

To avoid crossing of orders in our simulation models, we change the arrival moments of the outstanding lead times such that crossing does not occur, i.e. when there are outstanding orders at an ordering moment, a lead time is drawn independently from a fixed distribution. When this lead time crosses the lead times of outstanding orders, the arrival times of the replenishment orders are swapped (the initiation moment of the replenishments, however, are not swapped) such that the currently generated replenishment order will arrive after the outstanding replenishment orders have arrived. Hence, by rearranging the arrival times of the outstanding lead times, the expected length of the lead times will not change, but the variance will decrease. In a initialisation run, the variance of the non-crossing but dependent lead time process is calculated by simulation. This variance is then used to compute the relevant performance measures and control parameters. Note that for deterministic lead times or in case Q is large the previously discussed transformation has no effect.

2.1.3 Control policies

Apart from the usual control issues in an inventory system, i.e.

- how often should the inventory status be determined;
- when to place a replenishment order;
- how much to order.

we will pay in Chapters 7 and 8 also explicit attention to the question

- when to deliver the customer orders;
- from where to deliver the customer orders.

For the replenishment policy of an inventory system many strategies have been proposed. These strategies can be categorized in two groups.

- Continuous review policies are based on continuous observation of the system performance measures.
- Periodic review policies evaluate the system performance measures periodically at discrete points in time, after which possible replenishments are made.

For both situations one could make a distinction between fixed order sizes and variable order sizes. For the (s, Q) and (s, S) strategies a replenishment order is placed at the supplier whenever the inventory position (defined as the inventory on hand plus on order minus the backlog) drops below the reorder level s. Depending on the strategy the size of the order equals Q or S minus the current inventory position. A special case of the (s, S)

Table 2.1: Classification of inventory models

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

system is the base-stock system, where the reorder point equals $S-1$. This system is often denoted as the $(S-1, S)$ system, and is often used in service part inventory systems.

It has been shown that the (s, S) policy is optimal under general assumptions for the cost structure and the underlying stochastic processes, see Scarf (1960), Veinott (1966) and Hordijk and van der Duyn Schouten (1986). In spite of this result the (s, Q) policy is very popular in practice. Often there are additional constraints on the lot-size, for example, Q must be a multiple of a pallets size or some other package size. De Kok and Inderfurth (1997) discuss the nervousness in inventory management by comparing the (s, S), (s, Q) and (R, S) policies. In a multi-stage production environment variations in the (replenishment) order quantities can lead to nervousness in the planning system. Hence, when choosing the control policy it is also important to consider the setup stability and quantity stability of a control policy.

Under the regime of the periodic review policies the inventory position is monitored every R time units in order to take a replenishment decision. When the inventory position is below s, a replenishment order is placed at the supplier. Depending on the strategy the size of the order equals Q or S minus the current inventory position.

The main reason why periodic inventory models are used in practice lies in the advantages they have for both the supplier and the retailer. Using a periodic review replenishment policy (with a sufficiently large review period) for products ordered by the same supplier, the ordering and transportation costs can be reduced when replenishment orders for different products can be properly co-ordinated (see, for example, van Eijs (1993)). Secondly, for make-to-order organizations the knowledge of review moments of its customers (which are possible demand epochs), and the fact that the customers order only multiples of Q, can be translated into an efficient production schedule. This clearly reduces the production lead times, and consequently has a positive effect on the required inventory at the retailer to achieve a desired customer service level. Thus, the retailer benefits due to co-ordination and shorter lead times, whereas the manufacturer has more efficient production schedules and less work in progress. Furthermore, the (R, s, Q) or (R, s, S) inventory models coincide with the time phased reorder point in the MRP-system. MRP-systems are used within almost any manufacturing company.

2.1.4 System variables and performance measures

When demand is stochastic it is useful to categorize inventories as follows (see, Silver and Peterson (1985), Silver, Pyke and Peterson (1998))

- **Physical stock:** this stock is physically on the shelf. The physical stock level can never be negative. This system variable is relevant in determining whether a particular customer demand is satisfied directly from shelf or has to wait.

- **Net stock:** this stock equals the physical stock minus the backorders. The net stock is negative when there are backorders.

- **Inventory position:** equals the net stock plus stock on order. The stock on order is the stock which has been ordered but not yet received by the stock point. The inventory position is the key quantity in deciding whether or not to replenish.

- **Safety stock:** this stock is defined as the average level of the net stock just before a replenishment arrives. A positive safety stock provides a buffer against uncertainties in the lead time demand.

To evaluate the quality of the control of a inventory system, we introduce the most commonly used performance measures. We use the same notation as is used in Silver and Peterson (1985). Two groups of performance measures can be distinguished. The first group is related to the actual costs of an inventory system. We define

- B_1: the number of stockout occasions per unit of time. The cost of a stock-out is charged with a fixed value b_1, independent of the magnitude or duration of the stockout. A possible interpretation would be the costs of an expediting action.

- B_2: the average number of units short per unit of time. The costs per unit short are charged with a fixed value b_2, independent of the duration of the stockout. A situation where this type of costing would be appropriate is where units short are satisfied by overtime production or emergency orders.

- B_3: the average backlog level. The shortage cost per unit per unit of time are charged with a value b_3. An example of this type of costing would be a situation where the units under consideration are spare parts, each time unit short would result in a machine being idle.

- B_4: the average physical stock level. The holding costs per unit per unit of time are charged with a value b_4.

The second group of performance measures describes the level of service provided to customers. These measures are often used as constraints for the inventory system.

- P_1: the non-stockout probability per replenishment cycle. This service measure takes only the appearance of an backorder into consideration and not the magnitude or duration of the stock-out. This measure is also denoted as the α-service level (see, Schneider(1981)).

- P_2: the fraction of demand delivered directly from stock. This measure is very popular in practice (see, e.g., McLaughlin, Vastag, and Whybark (1994)). The P_2 service measure is also called the fill-rate or the β-service level (see, Schneider (1981)).
- P_3: the fraction of the time the net stock is positive. The P_3 service measure or ready rate finds common practice in case of equipment used for emergency purposes.
- P_4: the probability that an arbitrary customer has to wait. The waiting time of a customer is measured from the time of arrival of a customer until the time at which the demand is completely satisfied. This measure is often denote by $\mathbb{P}(W > 0)$.

2.2 Renewal theory background

In this section we provide a formal review of the basic concepts and results in renewal theory. Renewal theory concerns the study of stochastic processes counting the number of events that take place as a function of time. The inter occurrence times between events are independent and identically distributed (i.i.d.) random variables. Consider the sequence $\{X_n, n = 1, 2, \ldots\}$ of i.i.d. random variables. Let

$$F(x) = \mathbb{P}(X_n \leq x) \quad n = 1, 2, \ldots, x \geq 0, \tag{2.1}$$

denote the common cumulative distribution function (c.d.f.) of the sequence. Further assume that F is absolutely continuous with probability density function (p.d.f.) f. We denote $\mathbb{E}X_n^k$, $n, k = 1, 2, \ldots$, $\sigma^2_{X_n}$, and c_{X_n} as the k-the moment, the variance and the coefficient of variation of X_n respectively. Define

$$S_n = \sum_{k=1}^{n} X_k, \quad n = 1, 2, \ldots. \tag{2.2}$$

The c.d.f. and p.d.f. of S_n can be expressed in terms of the n-fold convolution of F and f. Thus we have

$$F^{n*}(x) = \int_0^x F^{(n-1)*}(x-y)dF(y) \quad n = 2, 3, \ldots, \tag{2.3}$$

$$f^{n*}(x) = \int_0^x f^{(n-1)*}(x-y)dF(y) \quad n = 2, 3, \ldots. \tag{2.4}$$

Let $N(t)$ denote the number of occurrences during $(0, t]$ where the origin coincides with an occurrence that is not counted. Since $\{S_n > t\} \Leftrightarrow \{N(t) < n\}$ it holds that for $n = 0, 1, \ldots$

$$\begin{aligned} \mathbb{P}(N(t) < n) &= 1 - F^{(n)*}(t), \\ \mathbb{P}(N(t) = n) &= F^{n*}(t) - F^{(n+1)*}(t). \end{aligned}$$

An important role in renewal theory is played by the renewal function $M(t)$ defined by

$$M(t) = \sum_{n=0}^{\infty} F^{n*}(t). \tag{2.5}$$

Thus we have

$$M(t) = \mathbb{E}N(t) + 1. \tag{2.6}$$

A useful characterisation of the renewal function is provided by the renewal equation

$$M(t) = F(t) + \int_0^t M(t-x)dF(x). \tag{2.7}$$

This result allows for the following important generalization. Let $a(t)$ be a given, integrable function that is bounded on finite intervals. Let $Z(t)$ satisfy

$$Z(t) = a(t) + \int_0^t Z(t-x)dF(x), \quad t \geq 0, \tag{2.8}$$

then this equation has a unique solution which is given by

$$Z(t) = \int_0^t a(t-x)dM(x). \tag{2.9}$$

For a proof we refer to Tijms (1994).

In many renewal theory related practical problems two interesting quantities appear quite often, namely the random variable U_t representing the time since the last occurrence time at time t (also known as the backward recurrence time) and the random variable V_t representing the time until the next occurrence time at time t (also known as the forward recurrence time). More precisely, U_t and V_t are given by

$$U_t = t - S_{N(t)}, \tag{2.10}$$

$$V_t = S_{N(t)+1} - t. \tag{2.11}$$

Denote by $v_t(.)$ the p.d.f. and by $V_t(.)$ the c.d.f. of V_t. The forward recurrence time V_t has a limiting distribution for $t \to \infty$.

Theorem 2.1 *Suppose $F(x)$ is non-arithmetic. Then*

$$\begin{aligned}
\lim_{t\to\infty} V_t(x) &= \frac{1}{\mathbb{E}X}\int_0^x (1-F(y))dy, x \geq 0 \\
\lim_{t\to\infty} v_t(x) &= \frac{1-F(x)}{\mathbb{E}X}, x \geq 0 \\
\lim_{t\to\infty} \mathbb{E}V_t^k &= \frac{\mathbb{E}X^{k+1}}{(k+1)\mathbb{E}X}, \quad k = 1, 2, \ldots
\end{aligned}$$

For the proof see, for example, Cox (1962) page 61. For U_t the same theorem holds.

Tijms (1994) shows by numerical investigations that the asymptotic results for the moments of V_t and U_t can be used when $t \geq Cond(X)$, where

$$Cond(X) = \begin{cases} \frac{3}{2}c_X^2 \mathbb{E}X & \text{if } c_X^2 > 1; \\ \mathbb{E}X & \text{if } 0.2 < c_X^2 \leq 1; \\ \frac{1}{2c_X}\mathbb{E}X & \text{if } 0 < c_X^2 \leq 0.2. \end{cases} \tag{2.12}$$

For derivation of several performance measures of the inventory models presented in the sequel of this thesis, the following lemma will be very useful.

Lemma 2.2 *Let M be the renewal function associated with F with mean μ, and let U be the associated asymptotic forward recurrence time distribution, then*

$$(M * U)(x) = \frac{x}{\mu} \tag{2.13}$$

Proof
Let $\tilde{F}(y)$ be the Laplace transform of F, thus $\tilde{F}(y) = \int_0^\infty e^{-yx}dF(x)$. Since $\tilde{U}(y) = (1 - \tilde{F}(y))/(y\mu)$ and $\tilde{M}(y) = 1/(1 - \tilde{F}(y))$, it follows that the Laplace transform of the convolution equals $1/(y\mu)$. Hence, taking the inverse Laplace transform of $1/(y\mu)$ yields $(M * U)(x) = x/\mu$. □

One of the basic problems in inventory theory is to determine the distribution of the demand during the lead time. In general this distribution function is intractable. Therefore we resort to its first three moments. Let the demand process be described by a compound renewal process. The sequences $\{A_n, n = 1, 2, \ldots\}$ and $\{D_n, n = 1, 2, \ldots\}$ represent the interarrival times and the demand sizes of customers arriving at the stock keeping unit, respectively. We denote by $N(0,t)$ the number of customer arrivals during $(0,t]$, and $D(0,t)$ represents the total demand during $(0,t]$. At time epoch zero the process is assumed to be stationary, i.e. the process has evolved over time infinitely long. We consider two situations

(i) The event stationary situation, i.e. at time zero an arrival occurs, in that case we write $N(0,t) \equiv N(t)$ and $D(0,t) \equiv D(t)$.

(ii) The time stationary situation, i.e. at time epoch zero is an arbitrary point in time. Since the process is stationary at zero, the c.d.f. of the first interarrival time after zero equals the c.d.f. of the asymptotic forward recurrence time associated to the sequence $\{A_n, n = 1, 2, \ldots\}$.

By using $D(0,t) = \sum_{n=1}^{N(0,t)} D_n$ and after some algebra, it can be shown that (see, for example, Sahin (1990) page 34),

$$\mathbb{E}D(0,t) = \mathbb{E}N(0,t)\mathbb{E}D \tag{2.14}$$

$$\begin{aligned} \mathbb{E}D(0,t)^2 &= \mathbb{E}N(0,t)\sigma^2(D) + \mathbb{E}N(0,t)^2(\mathbb{E}D)^2 && (2.15) \\ \mathbb{E}D(0,t)^3 &= \mathbb{E}N(0,t)\Big(\mathbb{E}D^3 - 3\mathbb{E}D^2\mathbb{E}D + 2(\mathbb{E}D)^3\Big) \\ &\quad + \mathbb{E}N(0,t)^2\Big(3\mathbb{E}D^2\mathbb{E}D - 3(\mathbb{E}D)^3\Big) + \mathbb{E}N(0,t)^3(\mathbb{E}D)^3 && (2.16) \end{aligned}$$

In order to come up with expressions for the moments of $\mathbb{E}N(0,t)$ we have to distinguish between the cases under assumption (i) and (ii). Let $\alpha_i^j := (\mathbb{E}A^i)^j$, then by using standard renewal theory, the following asymptotic relations can be derived (see, for example, Cox (1962)), when assumption (i) holds

$$\begin{aligned} \mathbb{E}N(t) &\simeq \frac{t}{\alpha_1^1} + \frac{\alpha_2^1}{2\alpha_1^2} - 1, && (2.17) \\ \mathbb{E}N(t)^2 &\simeq \frac{t^2}{\alpha_1^2} + t\left(\frac{2\alpha_2^1}{\alpha_1^3} - \frac{3}{\alpha_1^1}\right) + \frac{3\alpha_2^2}{2\alpha_1^4} - \frac{2\alpha_3^1}{3\alpha_1^3} - \frac{3\alpha_2^1}{2\alpha_1^2} + 1, && (2.18) \\ \mathbb{E}N(t)^3 &\simeq \frac{t^3}{\alpha_1^3} + t^2\left(\frac{9\alpha_2^1}{2\alpha_1^4} - \frac{6}{\alpha_1^2}\right) + t\left(\frac{9\alpha_2^2}{\alpha_1^5} - \frac{3\alpha_3^1}{\alpha_1^4} - \frac{12\alpha_2^1}{\alpha_1^3} + \frac{7}{\alpha_1^1}\right) && (2.19) \\ &\quad + \frac{3\alpha_4^1}{4\alpha_1^4} - \frac{6\alpha_2^1\alpha_3^1}{\alpha_1^5} + \frac{15\alpha_2^3}{2\alpha_1^6} + \frac{4\alpha_3^1}{\alpha_1^3} - \frac{9\alpha_2^2}{\alpha_1^4} + \frac{7\alpha_2^1}{2\alpha_1^2} - \alpha_1^1, \end{aligned}$$

and when assumption (ii) holds

$$\begin{aligned} \mathbb{E}N(0,t) &\simeq \frac{t}{\alpha_1^1}, && (2.20) \\ \mathbb{E}N(0,t)^2 &\simeq \frac{t^2}{\alpha_1^2} + t\left(\frac{\alpha_2^1}{\alpha_1^3} - \frac{1}{\alpha_1^1}\right) + \frac{\alpha_2^2}{2\alpha_1^4} - \frac{\alpha_3^1}{3\alpha_1^3}, && (2.21) \\ \mathbb{E}N(0,t)^3 &\simeq \frac{t^3}{\alpha_1^3} + t^2\left(\frac{3\alpha_2^1}{\alpha_1^4} - \frac{3}{\alpha_1^2}\right) + t\left(\frac{9\alpha_2^2}{2\alpha_1^5} - \frac{2\alpha_3^1}{\alpha_1^4} - \frac{3\alpha_2^1}{\alpha_1^3} + \frac{1}{\alpha_1^1}\right) && (2.22) \\ &\quad + \frac{\alpha_4^1}{2\alpha_1^4} - \frac{3\alpha_2^1\alpha_3^1}{\alpha_1^5} + \frac{3\alpha_2^3}{\alpha_1^6} + \frac{\alpha_3^1}{\alpha_1^3} - \frac{3\alpha_2^2}{2\alpha_1^4}. \end{aligned}$$

In the models of this thesis we will also encounter $N(L)$, where L denotes the lead time of replenishment orders. Often L is a random variable, independent of the renewal process $\{N(t)|t \geq 0\}$. In order to find approximations for the moments of $N(L)$ we can take expectations on the right hand side of (2.14) up to (2.22). However, in order to obtain reasonable approximations the following relation must hold: $\mathbb{P}(L < Cond(A)) \leq \epsilon$ for ϵ small enough. In case that this condition is not fulfilled, for a certain ϵ, the following approximation algorithm can be used for computing the moments of $N(L)$.

- Fit ME-distribution (see section 2.4) on $S_k = \sum_{j=1}^{k} A_j$, $k = 1, 2, \ldots$ and on L.
- Compute $\mathbb{P}(S_k \leq L)$, for which a closed form expression exists since S_k and L are both ME-distributed, for $k = 1, 2, \ldots$.

- Compute $P(N(L) = k) = I\!P(S_{k+1} \leq L) - I\!P(S_k \leq L)$ for $k = 1, 2, \ldots, k_{max}$, where k_{max} is chosen such that $\sum_{j=1}^{k_{max}} P(N(L) = j) \geq 0.99999$.

- Finally, compute the first two moments of $N(L)$ via

$$I\!EN(L)^m = \sum_{j=1}^{k_{max}} j^m I\!P(N(L) = j).$$

Note that this method strongly resembles the Gamma approximation suggested in Tijms (1994) page 16.

2.3 Relation between a service level and a cost perspective

Service level constraints are a well-established concept in inventory theory, and have been introduced as an alternative for stockout costs because of the intractability of the determination of shortage costs in most practical situations. The objective of inventory theory is to determine optimal inventory control policies. Optimality in itself, however, strongly depends on relevant criteria translated into objective functions and constraints. In the literature a distinction is made between objective functions that include shortage costs and objective functions that do not. In the latter case shortages are taken into account through the introduction of so-called service level constraints. In the first approach the expected sum of ordering, holding and shortage costs is minimized (see, e.g., Hadley and Whitin (1963) page 166, Silver and Peterson (1985) page 263, Federgruen and Zipkin (1984), Chen and Zheng(1993)), and this approach is often referred to as the "cost perspective". In case the shortage costs incorporate the customers' loss in goodwill and market share, then this problem can be seen as a long-term or strategic optimization problem.

In a practical situation it is often difficult to determine the value of the shortage costs. The second perspective circumvents this problem by minimizing the expected ordering plus holding costs subject to a service level constraint (see, e.g., Hadley and Whitin (1963) page 217, Tijms and Groenevelt (1984), Schneider and Rinquest (1990). This approach is often referred to as the "service level perspective". Since here we only consider the ordering and holding costs, and do not take into account the customers' loss in goodwill, we actually solve a short-term optimization problem. On the other hand, setting the appropriate value for the target service level is a long-term decision (this implicitly determines the customers' loss in goodwill and market share).

For both the definition of the shortage costs as well as for the service level many possibilities are available (see section 2.1.4).

When the function describing the shortage costs in the cost perspective equals the function describing the service level, then the optimal value for the Lagrange multiplier (or shadow price) in the constrained minimization problem can be interpreted as the appropriate value for the shortage costs. Consider, for example, the service level perspective with a P_2 service level constraint. Then the optimal value for the Lagrange multiplier (depending on the pre-defined target service level) represents the magnitude of the shortage cost in the cost-perspective based on the shortage costs related to the number of units short (i.e. the B_2 cost criterion). But, the value for the service-level-related shortage costs can be determined only by solving the constrained minimization problem, and therefore depends on the value for the optimal control parameters (see, for example, Aardal et al.(1989)).

An interesting question would be: is there a priori (without solving any optimization problem) an explicit relation between the value of the target service level and the value for the shortage costs? Silver and Peterson (1985), page 265, state that it can be shown that the use of the B_3 shortage costing measure leads to a decision rule equivalent to that for the P_2 service measure, where the equivalence is given by the relation $p_2 = b_3/(b_3 + b_4)$. Schneider (1981) states that such relation exists between the fixed backorder costs (B_1 cost criterion) and the fraction of time the net stock is positive (P_3 service level). For a mathematical proof of this last equivalence, Schneider referres to Klemm and Mikut (1972).

For reasons of illustration we now consider an (s, Q) inventory model. For the determination of the cost-optimal control parameters s and Q, we assume that the holding costs are proportional to the expected average physical stock level, i.e. to stock one unit of product costs b_4 per unit of time. The ordering costs are proportional to the expected number of replenishments per time unit. Placing an order costs a. Then define the following cost functions

$$C_1(s,Q;a,b_3,b_4) = \frac{a}{T(s,Q)} + b_3B_3(s,Q) + b_4B_4(s,Q), \tag{2.23}$$

$$C_2(s,Q;a,b_4) = \frac{a}{T(s,Q)} + b_4B_4(s,Q), \tag{2.24}$$

where $T(s,Q)$ is defined as the expected length of a replenishment cycle, and the $B_3(s,Q)$ and $B_4(s,Q)$ are given in section 2.1.4.

We consider the optimization and comparison of the following minimization problems.

$$\begin{array}{lll} (\mathcal{P}_1) & \text{minimize} & C_1(s,Q;a,b_3,b_4) \\ & \text{s.t.} & Q \geq 0; \\ (\mathcal{P}_2) & \text{minimize} & C_2(s,Q;a,b_4) \\ & \text{s.t.} & P_3(s,Q) = p_3,\ Q \geq 0; \\ (\mathcal{P}_3) & \text{minimize} & C_2(s,Q;a,b_4) \\ & \text{s.t.} & P_2(s,Q) = p_2,\ Q \geq 0. \end{array}$$

First we will show that a necessary condition for the optimal solution of $(\mathcal{P}_1)$ is given by $P_3(s,Q) = \frac{b_3}{b_3+b_4}$. Note the resemblance with the newsboy problem. This result will hold under the very general assumptions of the compound renewal demand process and stochastic lead times. This result was already derived for the (s,S) inventory model with a discrete time demand process and constant lead times (see Klemm (1974)).

Secondly, for the (s, Q) inventory model with a Poisson demand process it will be shown that the P_2 and P_3 service level are equal. This means that the statement in Silver and Peterson (1985), that a necessary condition for the unconstrained minimization problem with the B_3 cost criterion $(\mathcal{P}_1)$ is given by $P_2(s, Q) = \frac{b_3}{b_3+b_4}$, is only true for this kind of demand process, but not for the more general compound renewal process.

Seemingly, there is an explicit relation between the target P_3 service and the value of the shortage costs (b_3). This brings us to the final contribution of this section. We will show that solving $(\mathcal{P}_2)$ with $p_3 = \frac{b_3}{b_3+b_4}$ will in general not lead to the same solution as the unconstrained minimization problem $(\mathcal{P}_1)$. Furthermore, it follows from numerical examples that for situations where $b_3 \gg b_4$ the solutions actually do coincide.

The rest of this chapter is organized as follows. In section 2.3.1 the (s, Q) inventory model with a compound renewal demand process is defined in more detail. Relations are derived for the system variables and performance measures defined in section 2.1.4. Using these relations the necessary Kuhn-Tucker conditions are derived for $(\mathcal{P}_1)$ to $(\mathcal{P}_2)$ to compare the structures of the optimal solutions. Section 2.3.2 deals with some numerical examples that yield some intuitions for the magnitudes of the differences in the optimal control parameters for both perspectives.

2.3.1 The (s, Q) model description

In this single echelon (s, Q) inventory model we assume that the demand process is a compound renewal process. That is, the interarrival times of customers are described by the sequence $\{A_i\}_{i=1}^{\infty}$ of independent and identically distributed (i.i.d.) random variables with a common distribution function F_A, expectation $\mathbb{E}A$ and coefficient of variation c_A. A_i represents the time between the arrival of the i-th and $(i-1)$-th customer after time epoch 0. Further, let $\tilde{A}$ be the asymptotic forward recurrence time associated with the sequence $\{A_i\}_{i=1}^{\infty}$. The demand sizes of the customers are described by the sequence $\{D_i\}_{i=1}^{\infty}$ of i.i.d. random variables with a common distribution function F_D, expectation $\mathbb{E}D$ and coefficient of variation c_D. D_i represents the demand size of the i-th customer after time epoch 0. The sequence $\{D_i\}_{i=1}^{\infty}$ is independent of $\{A_i\}_{i=1}^{\infty}$.

Shortages are backordered, and replenishment decisions are based on the inventory position. The lead times are assumed to be i.i.d. random variables, with a common distribution function F_L. In spite of this assumption we will assume that deliveries of two successive replenishment orders do not cross in time (see section 2.1.3).

A well-known approach for deriving expressions for the previously defined performance measures is to focus on an arbitrary replenishment cycle. A replenishment cycle is defined as the time between two successive arrivals of replenishment orders (see, for example, Tijms and Groenevelt (1984)). The renewal reward theorem justifies the equality of the performance measures derived for that particular tagged replenishment cycle and the long-run performance measures (see, for example, Tijms (1994)).

Consider now an arbitrary replenishment cycle. Let time zero be an arbitrary customer arrival epoch. Denote the j-th ordering epoch after zero by σ_j. Let $D(t_1, t_2)$ be the total demand during $(t_1, t_2]$, and U_j the undershoot under s at σ_j, where U represents

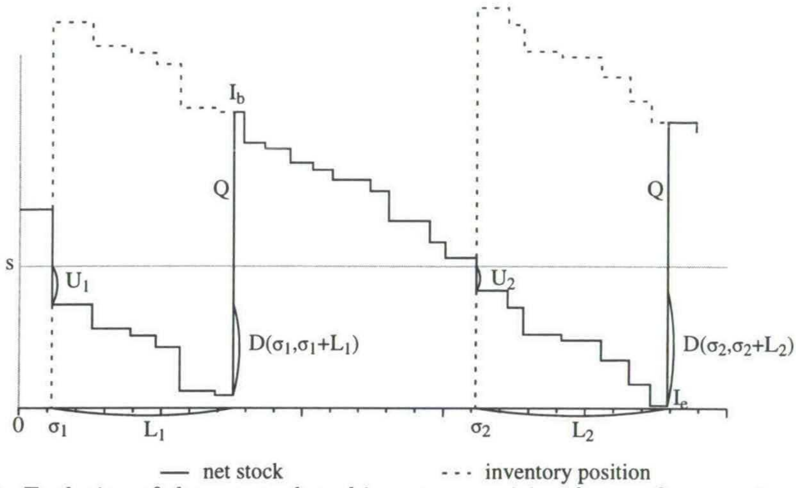


Figure 2.1: Evolution of the net stock and inventory position during the second replenishment cycle.

an arbitrary undershoot. L_j denotes the lead time of the j-th replenishment cycle after zero. Consider the second replenishment cycle after zero (see Figure 2.1). Denote, I_b as the net stock at the beginning of the second replenishment cycle after zero (just after the replenishment order has arrived), and I_e as the net stock at the end of the second replenishment cycle (just before the replenishment order arrives). Then it can be seen that (see Figure 2.1)

$$\begin{aligned} I_b &= s - U_1 + Q - D(\sigma_1, \sigma_1 + L_1) \overset{d}{=} \text{s+Q-Z}, \\ I_e &= s - U_2 - D(\sigma_2, \sigma_2 + L_2) \overset{d}{=} \text{s-Z}, \end{aligned}$$

where $\overset{d}{=}$ denotes equality in distribution and $Z = U + D(0, L)$.

Since demand is stationary and shortages are backordered, it can be shown that the folling theorem holds.

Theorem 2.3

$$T(s,Q) = \frac{Q\mathbb{E}A}{\mathbb{E}D}; \tag{2.25}$$

$$P_1(s,Q) = \mathbb{P}(I_e > 0); \tag{2.26}$$

$$P_2(s,Q) = 1 - \frac{\mathbb{E}(-I_e)^+ - \mathbb{E}(-I_b)^+}{Q}; \tag{2.27}$$

$$P_3(s,Q) = \gamma\frac{\mathbb{P}(I_b > 0) - \mathbb{P}(I_e > 0)}{Q} + \frac{\mathbb{E}(I_b + U)^+ - \mathbb{E}(I_e + U)^+}{Q}; \tag{2.28}$$

$$P_4(s,Q) \approx 1 - \frac{\mathbb{E}(D(0,L) + D - s)^+ - \mathbb{E}(D(0,L) + D - s - Q)^+}{Q} \tag{2.29}$$

$$B_1(s,Q) = \frac{\mathbb{E}D}{Q\mathbb{E}A}\mathbb{P}(I_e < 0); \tag{2.30}$$

ad (6) To derive the actual lead times from stockpoint 1 to stockpoint 2 the waiting time characteristics at stockpoint 1 are required. First consider constant transportation times l. Now we must distinguish between the situations $s \geq 0$ and $s < 0$.

Theorem 7.1 *Consider an (s, Q) inventory system with constant lead times l.*

$$\mathbb{P}(W \leq y) = 1 - \frac{\mathbb{E}(D(0, l-y) + D - s)^+ - \mathbb{E}(D(0, l-y) + D - s - Q)^+}{Q} \quad 0 \leq y < l,$$

$$\mathbb{P}(W = l) = \frac{\mathbb{E}(D - (s)^+)^+ - \mathbb{E}(D - (s)^+ - Q)^+}{Q},$$

and

$$\mathbb{P}(l < W \leq y) = \frac{\mathbb{E}D(y-l) - \mathbb{E}(D(y-l) + s)^+}{Q} \quad y > l, s < 0.$$

For the proof see Appendix 7.B.

By using Theorem 7.1 the first two moments of W can be determined. First consider $s \geq 0$. Using Theorem 7.1 yields

$$\begin{aligned}
\mathbb{E}W &= \int_0^\infty \mathbb{P}(W > y) dy \\
&= \int_0^l \frac{\mathbb{E}(D(0, l-y) + D - s - Q)^+ - \mathbb{E}(D(0, l-y) + D - s)^+}{Q} dy \\
&= l \int_0^l \frac{\mathbb{E}(D(0, y) + D - s - Q)^+ - \mathbb{E}(D(0, y) + D - s)^+}{Q} 1/l\, dy \\
&= l \frac{\mathbb{E}(D(0, \tilde{L}) + D - s - Q)^+ - \mathbb{E}(D(0, \tilde{L}) + D - s)^+}{Q}, \qquad (7.22)
\end{aligned}$$

where $\tilde{L}$ is uniformly distributed over $(0, l)$. Analogously, we find

$$\mathbb{E}W^2 = l^2 \frac{\mathbb{E}(D(0, \hat{L}) + D - s - Q)^+ - \mathbb{E}(D(0, \hat{L}) + D - s)^+}{Q}, \qquad (7.23)$$

where

$$F_{\hat{L}}(t) = \begin{cases} 0 & t < 0; \\ \int_0^t \frac{2(l-y)}{l^2} dy & t < l; \\ 1 & t > l. \end{cases}$$

Note that both $\tilde{L}$ as $\hat{L}$ have substantial probability mass near zero. Therefore we have to be careful with applying (2.14), (2.15) with the asymptotic expressions (2.20) and (2.21) zero is an arbitrary moment in time). By conditioning on the first arriving customer after time epoch 0, we find

$$H(x) = x\mathbb{E}A + \int_0^x H(x-y)dF_D(y). \tag{2.34}$$

$$T(x) = \mathbb{E}A + \int_0^x T(x-y)dF_D(y). \tag{2.35}$$

Let M be the renewal function associated with F_D, then solving (2.34) and (2.35) yields

$$H(x) = \mathbb{E}A \int_0^x (x-y)dM(y) \tag{2.36}$$

$$T(x) = \mathbb{E}A\, M(x) \tag{2.37}$$

Consider the situation that zero is an arbitrary point in time, and let A_0 be the arrival time of the first customer after zero. Then A_0 equals $\tilde{A}$ (the forwards recurrence time with respect to the arrival process of customers). By conditioning on the first arriving customer after time epoch zero results into

$$\tilde{H}(x) = x\mathbb{E}\tilde{A} + \int_0^x H(x-y)dF_D(y). \tag{2.38}$$

$$\tilde{T}(x) = \mathbb{E}\tilde{A} + \int_0^x T(x-y)dF_D(y). \tag{2.39}$$

Combining these result with relations (2.34) to (2.37) yields

$$\tilde{H}(x) = (\mathbb{E}\tilde{A} - \mathbb{E}A)x + \mathbb{E}A \int_0^x (x-y)dM(y) \tag{2.40}$$

$$\tilde{T}(x) = (\mathbb{E}\tilde{A} - \mathbb{E}A) + \mathbb{E}A\, M(x) \tag{2.41}$$

Now, consider an arbitrary replenishment cycle. Recall, that I_b denotes the net stock at the beginning of a replenishment cycle (just after a replenishment order has arrived), and I_e denotes the net stock at the end of the replenishment cycle (just before a replenishment order has arrived). Then the expected area between the physical inventory level and the zero level from the beginning of the arbitrary replenishment cycle until infinity is given by $\mathbb{E}\tilde{H}(I_b^+)$. Subtracting the expected area between the physical inventory level and the zero level from the end of the arbitrary replenishment cycle until infinity, which is given by $\mathbb{E}\tilde{H}(I_e^+)$, yields the expected area between the physical inventory level and the zero level during the arbitrary replenishment cycle. The same sort of reasoning holds for the expected time the net stock is positive. Dividing by the expected length of the replenishment cycle yields

$$P_3(s,Q) = \frac{\mathbb{E}\tilde{T}(I_b^+) - \mathbb{E}\tilde{T}(I_e^+)}{T(s,Q)}; \tag{2.42}$$

$$B_4(s,Q) = \frac{\mathbb{E}\tilde{H}(I_b^+) - \mathbb{E}\tilde{H}(I_e^+)}{T(s,Q)}. \tag{2.43}$$

By using Lemma 2.2 we can rewrite $\mathbb{E}\tilde{H}(I_b^+)$

$$\begin{aligned}
\mathbb{E}\tilde{H}(I_b^+) &= \int_0^{s+Q} \tilde{H}(s+Q-x)dF_{D(0,L)+U}(x) \\
&\approx (\mathbb{E}\tilde{A} - \mathbb{E}A) \int_0^{s+Q} (s+Q-x)dF_{D(0,L)+U}(x) \\
&\quad + \frac{\mathbb{E}A}{2\mathbb{E}D} \int_0^{s+Q} (s+Q-x)^2 dF_{D(0,L)}(x) \\
&= (\mathbb{E}\tilde{A} - \mathbb{E}A)\mathbb{E}(I_b^+) + \frac{\mathbb{E}A\mathbb{E}((I_b+U)^+)^2}{2\mathbb{E}D}.
\end{aligned} \tag{2.44}$$

The approximation is due to the fact that we approximate the undershoot by the asymptotic forward recurrence time in order to use Lemma 2.2. For the exact c.d.f. of the undershoot we refer to de Kok (1991b). We can rewrite $\mathbb{E}\tilde{H}(I_e^+)$, $\mathbb{E}\tilde{T}(I_b^+)$, and $\mathbb{E}\tilde{T}(I_e^+)$ in the same way.

Finally substituting (2.25) and (2.44) in (2.43) and (2.42) yields (2.28) and (2.33), respectively. For the proof of expression (2.32) for the expected average backlog we will use the well-known relation that the inventory position equals the physical stock plus on order minus the backlog (see, for example, Hadley and Whitin (1963) page 187). Taking expectations on both sides of the equation and using that the expected inventory position is equal to $s + Q/2$, and that the expected amount on order is given by $\frac{\boldsymbol{EDEL}}{\boldsymbol{EA}}$, we find (2.32) which completes the proof $\square$.

First we consider $(\mathcal{P}_1)$. Since $a > 0$, it can be shown that the optimal value of Q does not lie on the boundary of the feasible region $Q \geq 0$. Hence, the necessary conditions for s^* and Q^* are given by:

$$\frac{\partial C_1(s,Q;a,b_3,b_4)}{\partial s} = \frac{\partial C_1(s,Q;a,b_3,b_4)}{\partial Q} = 0. \tag{2.45}$$

Substituting relations (2.25), (2.33) and (2.32) into definition (2.23) and using conditions (2.45) yields,

$$P_3(s,Q) = \frac{b_3}{b_3+b_4}; \tag{2.46}$$

$$(b_3+b_4)\Big(\gamma\frac{\mathbb{P}(I_b>0)}{Q} + \frac{\mathbb{E}(I_b+U)^+}{Q} - \frac{B_4(s,Q)}{Q}\Big) = \frac{a\mathbb{E}D}{Q^2\,\mathbb{E}A} + \frac{1}{2}b_3. \tag{2.47}$$

In Appendix 2.A we also give an intuitive proof for relation (2.46). This alternative proof holds for very general demand processes and control policies.

For the situation that $s+Q$ is sufficiently large $\mathbb{P}(I_b > 0) = 1$, $\mathbb{E}(I_b+U)^+ = s+Q-\mathbb{E}D(\sigma_1, \sigma_1+L_1)$, and some terms in $B_4(s,Q)$ can be elaborated further. Hence, we can rewrite (2.47) into

$$Q^2 = \frac{b_4+b_3}{b_4}\Big(\mathbb{E}(-(I_e+U)^+)^2 - \gamma\mathbb{E}(-(I_e)^+)\Big) + \frac{2a\mathbb{E}D}{b_4\mathbb{E}A}. \tag{2.48}$$

This approximation is often used (see, for example, formula (4-89) in Hadley and Whitin (1963)). However, Zipkin (1986b) argued that these approximations can lead to poor results.

Of course we can solve s^* and Q^* from (2.46) and (2.47). Alternatively, we may use the simple relation (2.46) to obtain the optimal reorder point for a given value of Q (denoted by $s^*(Q)$). Provided that $C_1(s^*(Q), Q; a, b_3, b_4)$ is convex in Q, the remaining one-dimensional optimization problem can easily be solved for example by Golden Section search. Conditions under which the loss-function (sum of holding and backordering costs) is quasiconvex are derived in Chen and Zheng (1993).

Note that in case U is negligible we can write $\mathbb{E}(I_b+U)^+ - \mathbb{E}(I_e+U)^+$ as $Q - \mathbb{E}(-I_e)^+ + \mathbb{E}(-I_b)^+$. When the arrival process is Poisson then $\tilde{A} \overset{d}{=} A$, hence $\gamma = 0$, where $X_1 \overset{d}{=} X_2 \Leftrightarrow \forall x \mathbb{P}(X_1 \leq x) = \mathbb{P}(X_2 \leq x)$.

So for the situation where U is negligible and Poisson arrivals $P_2(s,Q)$ equals $P_3(s,Q)$. We now consider the two constrained optimization problems $(\mathcal{P}_2)$ and $(\mathcal{P}_3)$. Define for $i = 1, \ldots, 3$, (s_i^*, Q_i^*) as the solution of $(\mathcal{P}_i)$. As was indicated in the introduction Silver and Peterson (1985) stated that, provided that $p_2 = \frac{b_3}{b_3+b_4}$, (s_1^*, Q_1^*) is equal to (s_2^*, Q_2^*). Relation (2.46), however, shows that (s_1^*, Q_1^*) must satisfy $P_3(s_1^*, Q_1^*) = \frac{b_3}{b_3+b_4}$, and not $P_2(s_1^*, Q_1^*) = \frac{b_3}{b_3+b_4}$.

Now we know that (s_1^*, Q_1^*) must satisfy $P_3(s_1^*, Q_1^*) = \frac{b_3}{b_3+b_4}$ we focus on problem $(\mathcal{P}_2)$. What remains is the question whether (s_1^*, Q_1^*) and (s_2^*, Q_2^*) are the same. From the theory of Lagrange multipliers we know that $(\mathcal{P}_2)$ is equivalent to

$$\text{minimize}\{\mathcal{L}_3(s, Q, \lambda) = C_2(s, Q; a, b_4) + \lambda(P_3(s, Q) - p_3)|Q \geq 0\} \tag{2.49}$$

Again, using that $Q^* > 0$ and the complementary slackness condition, the necessary Kuhn-Tucker conditions state that

$$\begin{aligned} P_3(s, Q) &= p_3, & (2.50)\\ b_4 P_3(s, Q) &= \lambda \frac{\partial P_3(s, Q)}{\partial s}, & (2.51)\\ b_4\left(\frac{\gamma \mathbb{P}(I_b > 0) + \mathbb{E}(I_b + U)^+ - B_4(s, Q)}{Q}\right) &= \frac{a\mathbb{E}D}{Q^2\, \mathbb{E}A} - \lambda \frac{\partial P_3(s, Q)}{\partial Q}. & (2.52) \end{aligned}$$

Eliminating λ from (2.51) and (2.52) yields

$$\begin{aligned} P_3(s, Q) &= p_3, & (2.53)\\ b_4\left(\frac{\gamma \mathbb{P}(I_b > 0) + \mathbb{E}(I_b + U)^+ - B_4(s, Q)}{Q}\right) &= \frac{a\mathbb{E}D}{Q^2\, \mathbb{E}A} - b_4 P_3(s, Q) \frac{\frac{\partial P_3(s,Q)}{\partial Q}}{\frac{\partial P_3(s,Q)}{\partial s}}. & (2.54) \end{aligned}$$

Comparing (s_1^*, Q_1^*) with (s_2^*, Q_2^*) it is evidently true that both solution only can coincide when $p_3 = \frac{b_3}{b_3+b_4}$. However, in general equations (2.47) and (2.54) do not lead to the same solution. Therefore, we may conclude that in general solving $(\mathcal{P}_1)$ or $(\mathcal{P}_2)$ will not lead to the same solutions, not even when $p_3 = \frac{b_3}{b_3+b_4}$. In the numerical examples of section 2.3.2 indeed this inequality is shown. The difference between (s_1^*, Q_1^*) and (s_2^*, Q_2^*) turn out to be very large in some situations, but when $b_3 \gg b_4$ the differences are small.

2.3.2 Numerical results

In this section we compare for $i = 1, 2, 3$, (s_i^*, Q_i^*) for various values of the input parameters. For the input parameters we distinguish between the cost parameters (a, b_3, b_4), service level targets (p_2, p_3), and system parameters defining the underlying demand and lead time processes $(\mathbb{E}D, c_D, \mathbb{E}A, c_A, \mathbb{E}L, c_L)$.

We assume that $\tilde{A}$ can be approximated by it's asymptotic relations (see theorem (2.1),

$$\begin{aligned} \mathbb{E}\tilde{A} &\approx \frac{\mathbb{E}A^2}{2\mathbb{E}A}, & (2.55)\\ \mathbb{E}\tilde{A}^2 &\approx \frac{\mathbb{E}A^3}{3\mathbb{E}A}. & (2.56) \end{aligned}$$

We assume that a time unit is equal to 1 day, and that a year equals 250 days. The service level parameters are always chosen such that $p_2 = p_3 = \frac{b_3}{b_3+b_4}$. In the first numerical experiment we compare (s_1^*, Q_1^*) and (s_2^*, Q_2^*). The system parameters are fixed whereas the cost parameters are varied. In Table 2.2 and Table 2.3 the results are shown. In these tables we present (s_1^*, Q_1^*), (s_2^*, Q_2^*) and the annual total relevant costs associated with the optimal control parameters (where, $C_{i,j} = 250\, C_i(s_j^*, Q_j^*)$ for $i \in \{1,2\}, j \in \{1,2,3\}$).

From these results we indeed conclude that in general (s_1^*, Q_1^*) and (s_2^*, Q_2^*), $C_{1,1}$ and $C_{1,2}$, and $C_{2,1}$ and $C_{2,2}$ do not coincide. However, for values of $b_3 \gg b_4$, i.e. a high service level p_3, we see that the solutions actually are the same. This hypothesis has been tested and confirmed for quite a number of different values for the system parameters.

Table 2.2: The comparison of (s_1^*, Q_1^*) and (s_2^*, Q_2^*) where $b_4 = 0.01$, $\mathbb{E}D = 5$, $c_D = 0.1$, $\mathbb{E}A = 1$, $c_A = 0.1$, $\mathbb{E}L = 10$, $c_L = 0.1$

		$(\mathcal{P}_1)$				$(\mathcal{P}_2)$			
a	p_3	s_1^*	Q_1^*	$C_{1,1}$	$C_{2,1}$	s_2^*	Q_2^*	$C_{1,2}$	$C_{2,2}$
0.1	0.5000	41.0	17.8	23.2	15.0	38.6	22.7	23.8	14.6
0.1	0.9091	53.0	14.0	42.8	34.8	52.8	14.8	42.9	34.7
0.1	0.9524	55.7	13.5	48.2	40.7	55.5	14.1	48.3	40.7
0.1	0.9901	60.9	12.9	59.5	53.0	60.8	13.2	59.5	53.0
10.0	0.5000	-21.0	142.0	177.5	132.8	-50.3	200.5	188.1	125.2
10.0	0.9091	40.4	106.9	243.3	228.1	40.0	110.8	243.4	227.9
10.0	0.9524	45.9	105.1	252.6	240.6	45.8	106.8	252.6	240.6
10.0	0.9901	53.9	103.5	268.5	260.1	53.9	104.1	268.5	260.1

Table 2.3: The comparison of (s_1^*, Q_1^*) and (s_2^*, Q_2^*) where $b_4 = 0.01$, $\mathbb{E}D = 5$, $c_D = 2$, $\mathbb{E}A = 1$, $c_A = 2$, $\mathbb{E}L = 2$, $c_L = 0.5$

		$(\mathcal{P}_1)$				$(\mathcal{P}_2)$			
a	p_3	s_1^*	Q_1^*	$C_{1,1}$	$C_{2,1}$	s_2^*	Q_2^*	$C_{1,2}$	$C_{2,2}$
5.0	0.5000	-43.4	103.7	131.8	95.0	-62.5	142.7	138.3	90.5
5.0	0.9091	0.7	93.0	219.4	167.3	1.9	84.3	220.2	166.4
5.0	0.9524	10.7	99.2	260.9	193.9	11.9	93.3	261.3	193.5
5.0	0.9901	54.2	103.0	379.9	302.5	54.4	102.2	379.9	302.5
10.0	0.5000	-63.6	144.8	182.1	133.7	-91.8	201.8	192.1	126.8
10.0	0.9091	-1.8	120.7	277.7	231.3	-1.0	110.8	278.7	230.5
10.0	0.9524	6.1	127.8	315.8	252.7	7.2	119.6	316.4	252.1
10.0	0.9901	47.1	133.3	432.7	355.7	47.3	132.3	432.7	355.6

Finally, we want to compare (s_2^*, Q_2^*) and (s_3^*, Q_3^*). In case the undershoot is negligible and in case of Poisson arrivals, expressions $P_2(s,Q)$ equals $P_3(s,Q)$. Therefore, we only considered situations where these conditions are not fulfilled, i.e., when $\mathbb{E}D = 5$, $c_A = 0.1$ or $c_A = 2$ (smooth or erratic arrival patterns of the customers). The target services p_2

and p_3 where chosen to be equal, and are varied between 0.60 and 0.9999. The results are tabulated in Table 2.4. These results show that especially for erratic demand and lead time processes (s_2^*, Q_2^*) and (s_3^*, Q_3^*) differ. But for smooth processes and relative high service levels $(\mathcal{P}_2)$ and $(\mathcal{P}_3)$ yield the same results. Furthermore, we see that $C_{2,2}$ and $C_{2,3}$ differ only very little in all situations.

Table 2.4: The comparison of (s_2^*, Q_2^*) and (s_3^*, Q_3^*) where $a = 0.1$, $b_4 = 0.001$, $\mathbb{E}D = 5$, $\mathbb{E}A = 1$, $\mathbb{E}L = 5$.

	$c_D = 0.25, c_A = 0.1, c_L = 0.1$						$c_D = 2, c_A = 2, c_L = 1$					
	$(\mathcal{P}_3)$			$(\mathcal{P}_2)$			$(\mathcal{P}_3)$			$(\mathcal{P}_2)$		
$p_2 = p_3$	s_3^*	Q_3^*	$C_{2,3}$	s_2^*	Q_2^*	$C_{2,2}$	s_3^*	Q_3^*	$C_{2,3}$	s_2^*	Q_2^*	$C_{2,2}$
0.6000	0.0	60.5	4.7	3.7	53.3	4.8	10.1	63.4	9.0	-5.9	51.6	5.4
0.6250	6.0	51.2	5.0	5.8	51.1	5.0	12.7	63.5	9.6	-3.8	49.5	5.6
0.6500	8.0	49.2	5.2	7.8	49.2	5.2	15.4	63.5	10.2	-1.9	47.8	5.8
0.6750	9.8	47.4	5.4	9.5	47.6	5.4	18.5	63.4	10.8	-0.1	45.9	6.1
0.7000	11.4	45.9	5.6	11.2	45.9	5.6	21.7	63.4	11.5	0.7	49.1	6.3
0.7250	13.0	44.1	5.8	12.9	44.1	5.8	25.2	63.4	12.3	2.6	49.8	6.7
0.7500	14.5	42.6	6.0	14.4	42.6	6.0	29.0	63.4	13.1	4.7	50.6	7.1
0.7750	15.9	41.4	6.2	15.7	41.4	6.2	33.2	63.4	14.1	7.1	52.0	7.7
0.8000	17.2	40.2	6.4	17.0	40.3	6.4	37.8	63.5	15.1	10.2	53.0	8.3
0.8250	18.4	39.2	6.7	18.2	39.5	6.6	43.2	63.2	16.3	13.7	55.0	9.2
0.8500	19.6	38.3	6.9	19.4	38.7	6.9	49.3	63.3	17.8	18.3	57.0	10.3
0.8750	20.8	37.5	7.1	20.7	37.8	7.1	56.6	62.7	19.4	24.2	59.3	11.7
0.9000	22.1	36.8	7.4	22.1	37.2	7.4	65.3	62.7	21.5	32.1	61.6	13.7
0.9250	23.5	36.2	7.7	23.5	36.5	7.7	76.5	62.7	24.2	43.5	62.8	16.4
0.9500	25.2	35.4	8.0	25.3	35.7	8.1	92.2	62.5	28.0	60.3	63.8	20.4
0.9750	27.5	34.9	8.6	27.9	34.9	8.7	118.6	62.4	34.4	89.9	64.3	27.6
0.9900	30.2	34.4	9.2	30.6	34.4	9.3	153.2	62.0	43.0	129.5	64.7	37.3
0.9990	35.5	33.7	10.5	36.1	33.8	10.7	238.5	61.4	64.2	229.6	64.7	62.3
0.9999	39.9	33.6	11.6	40.6	33.6	11.8	322.1	61.3	85.1	329.2	64.8	87.2

2.4 Fitting a mixed Erlang distribution based on the first two moments

All the inventory models in this thesis make use of stochastic variables. To verify or implement the inventory models we need to specify the distribution function of these stochastic variables. A versatile class of distribution functions is the class of mixtures of two Erlang distributions (denoted by ME distributions). When using ME distributions we often can explore the models further, whereas most of the generality remains.

Let X be a ME-distributed random variable with c.d.f. F, p.d.f. f and mean $\mathbb{E}X$ and coefficient of variation c_X. The density function f of X is described by

$$f(x) = \sum_{j=1}^{2} p_j \mu_j^{k_j} \frac{x^{k_j-1}}{(k_j-1)!} e^{-\mu_j x}, \quad x \geq 0, \tag{2.57}$$

where $p_1 \geq 0$, $p_2 \geq 0$, $p_1 + p_2 = 1$, $k_1, k_2 \in \mathbb{N}$. Since $p_1 + p_2 = 1$ we have 5 degrees of freedom.

In Tijms (1994) page 358, and van der Heijden (1993) algorithms are described to determine the values of $p_1, p_2, k_1, k_2, \mu_1$, and μ_2, based on the first two moments. When the coefficient of variation is smaller than 1, the following density function is advocated

$$f(x) = p_1 \mu_1^{k-1} \frac{x^{(k-2)}}{(k-2)!} e^{-\mu_1 x} + (1-p_1)\mu_2^k \frac{x^{(k-1)}}{(k-1)!} e^{-\mu_2 x} \quad x \geq 0. \tag{2.58}$$

Hence, $p_2 = (1-p_1), k_1 = k-1, k_2 = k$. The other parameters satisfy the following equations:

$$\begin{aligned}
k &= \lfloor \frac{1}{c_X^2} + 1 \rfloor; \\
p_1 &= \frac{1}{1+c_X^2}\left(kc_X^2 - \sqrt{k(1+c_X^2) - k^2 c_X^2}\right); \\
\mu_1 &= \frac{k-p_1}{\mathbb{E}X}; \\
\mu_2 &= \mu_1.
\end{aligned}$$

When the coefficient of variation is larger or equal to 1, the following density function is advocated

$$f(x) = p_1 \mu_1 e^{-\mu_1 x} + (1-p_1)\mu_2 e^{-\mu_2 x} \quad x \geq 0. \tag{2.59}$$

Hence, $p_2 = (1-p_1), k_1 = k_2 = 1$. The other parameters satisfy the following equations:

$$\begin{aligned}
\mu_1 &= \frac{2}{\mathbb{E}X}\left(1 + \sqrt{\frac{c_X^2 - \frac{1}{2}}{c_X^2 + 1}}\right); \\
\mu_2 &= \frac{4}{\mathbb{E}X} - \mu_1; \\
p_1 &= \frac{\mu_1(\mu_2 \mathbb{E}X - 1)}{\mu_2 - \mu_1}.
\end{aligned}$$

Let $n, m \in \mathbb{N}$, $z \in \mathbb{R}^+$ and let Z be a ME-distributed random variable with c.d.f. G, p.d.f. g and mean $\mathbb{E}Z$ and coefficient of variation c_Z. The density function of Z is give by

$$g(z) = \sum_{j=1}^{2} q_j \rho_j^{l_j} \frac{z^{l_j-1}}{(l_j-1)!} e^{-\rho_j z}, \quad z \geq 0, \tag{2.60}$$

Furthermore we define the following two auxiliary functions

$$
\begin{aligned}
a_1(n,z) &= \sum_{j=0}^{n-1} \frac{z^j}{j!} e^{-z} \\
a_2(m,k,n,l,z) &= \sum_{j=0}^{m+k} \frac{(j+l+n)!}{j!l!} z^j
\end{aligned}
$$

An important feature of ME distributions is that closed form expressions for partial moments, such as $\int_z^\infty x dF(x)$ exist. In view of computational aspects of the inventory models presented in this thesis the following functions turn out to be very useful

$$
\begin{aligned}
b_1(n,m,z) &= z^n \int_z^\infty x^m dF(x); \\
b_2(n,m,z) &= z^n \int_0^z x^m dF(x); \\
b_3(n,m) &= \int_0^\infty z^n \int_z^\infty x^m dF(x) dG(z); \\
b_4(n,m) &= \int_0^\infty z^n \int_0^z x^m dF(x) dG(z).
\end{aligned}
$$

The implementation of these methods are of course dependent on the values of n, m and z. We resort to the derivation of the most complex situations.
If $z > 0, n > 0, m \geq 0$ then

$$
\begin{aligned}
b_1(n,m,z) &= \int_z^\infty z^n x^m dF(x) \\
&= z^n \sum_{j=1}^{2} p_j \int_z^\infty x^m \mu_j^{k_j} \frac{x^{(k_j-1)}}{(k_j-1)!} e^{-\mu_j x} dx \\
&= z^n \sum_{j=1}^{2} p_j \frac{(k_j+m-1)!}{\mu_j^m (k_j-1)!)} \int_z^\infty \mu_j^{(k_j+m)} \frac{x^{(k_j+m-1)}}{(k_j+m-1)!} e^{-\mu_j x} dx \\
&= z^n \sum_{j=1}^{2} p_j \frac{(k_j+m-1)!}{\mu_j^m (k_j-1)!)} \sum_{i=0}^{k_j+m-1} \frac{(\mu_j z)^i}{i!} e^{-\mu_j z} \\
&= z^n \sum_{j=1}^{2} p_j \frac{(k_j+m-1)!}{\mu_j^m (k_j-1)!)} a_1(m+k_j, \mu_j z).
\end{aligned}
$$

And for $n > 0, m \geq 0$

$$
b_3(n,m) = \int_0^\infty \int_z^\infty z^n x^m dF(x) dG(z)
$$

$$
\begin{aligned}
&= \sum_{j=1}^{2}\sum_{i=1}^{2} p_j q_i \int_0^\infty z^n \frac{(m+k_j-1)!}{\mu_j^m (k_j-1)!} \sum_{t=0}^{k_j+m-1} \frac{(\mu_j z)^t}{t!} e^{-\mu_j z} \rho_i^{l_i} \frac{z^{(l_i-1)}}{(l_i-1)!} e^{-\rho_i z} dz \\
&= \sum_{j=1}^{2}\sum_{i=1}^{2} p_j q_i \frac{(m+k_j-1)!}{\mu_j^m (k_j-1)!} \sum_{t=0}^{k_j+m-1} \frac{(t+l_i+n-1)!}{t!(l_i-1)!} \frac{\mu_j^t \rho_i^{l_i}}{(\mu_j+\rho_i)^{t+l_i+n}} \\
&= \sum_{j=1}^{2}\sum_{i=1}^{2} p_j q_i \frac{(m+k_j-1)!}{\mu_j^m (k_j-1)!} \frac{\rho_i^{l_i}}{(\mu_j+\rho_i)^{l_i+n}} a_2(m, k_j-1, n, l_i-1, (\mu_j/(\mu_j+\rho_i)))
\end{aligned}
$$

These functions can be used efficiently to design other interesting quantities such as $\mathbb{E}(X-z)^+, \mathbb{E}((X-Z)^+)^2$. We now give a list of functions which are often encountered in inventory models

$$
\begin{aligned}
\mathbb{E}((X-z)^+)^m &= \sum_{k=0}^{m} \binom{m}{m-k} (-1)^k b_1(k, m-k, z), \\
\mathbb{E}((z-X)^+)^m &= \sum_{k=0}^{m} \binom{m}{m-k} (-1)^k b_2(k, m-k, z), \\
\mathbb{E}((X-Z)^+)^m &= \sum_{k=0}^{m} \binom{m}{m-k} (-1)^k b_3(k, m-k), \\
\mathbb{E}((Z-X)^+)^m &= \sum_{k=0}^{m} \binom{m}{m-k} (-1)^k b_4(k, m-k), \\
\mathbb{E}(\max\{X, z\})^m &= b_1(0, m, z) + b_2(m, 0, z), \\
\mathbb{E}(\max\{X, Z\})^m &= b_3(0, m) + b_4(m, 0), \\
\mathbb{E}(\min\{X, z\})^m &= b_1(m, 0, z) + b_2(0, m, z), \\
\mathbb{E}(\min\{X, Z\})^m &= b_3(m, 0) + b_4(0, m).
\end{aligned}
$$

Example: Let Z denote the random variable representing the sum of the lead time demand plus the undershoot. The P_2 service for the (s, Q) system is given by

$$
P_2(s, Q) = 1 - \frac{\mathbb{E}(Z-s)^+ - \mathbb{E}(Z-(s+Q))^+}{Q},
$$

see (2.27). Now using that Z is ME-distributed, we simply can use the algebraic form above to implement this P_2 service measure, namely

$$
P_2(s, Q) = 1 - (b_1(0, 1, s) - b_1(1, 0, s) - b_1(0, 1, s+Q) + b_1(1, 0, s+Q))/Q.
$$

Appendix 2.A Alternative proof for (2.46)

Consider an inventory system with a stationary demand process. The replenishments are controlled by a reorder level s, and shortages are backordered.

Now we focus on the B_3, B_4 and P_3 performance measures in an arbitrary replenishment cycle, see Figure 2.2 When we consider a marginal change Δs of the reorder level s. The

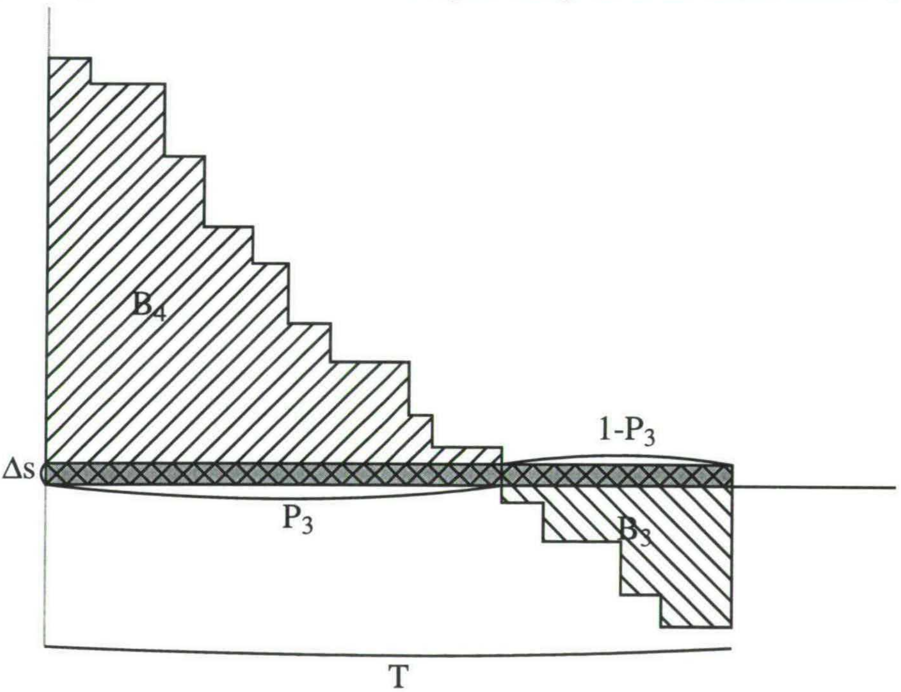


Figure 2.2: Arbitrary replenishment cycle of the inventory system

change in holding costs is given by $\frac{\Delta s\, b_4\, P_3}{T}$ and the change in shortage costs is given by $\frac{\Delta s\, b_3\, (1-P_3)}{T}$. Hence the optimal s must satisfy

$$\frac{\Delta s\; b_4\; P_3}{T} = \frac{\Delta s\; b_3\; (1 - P_3)}{T}$$

which yields

$$P_3 = \frac{b_3}{b_3 + b_4}$$

□

Chapter 3

The (R, s, Q) inventory model with Compound Bernoulli demand

This chapter is based on Janssen, Heuts and de Kok (1998). The (R, s, Q) inventory model has been studied extensively during the last decades. Under the regime of this inventory policy, every R time units the inventory position is monitored in order to make a replenishment decision. When the inventory position is below s, a multiple of Q is ordered such that the inventory position is raised to a value between s and $s+Q$ (see section 2.1.4).

Many heuristic and optimal methods are developed to determine the values of the control parameters: R, s and Q (see section 2.3). Here we focus on a service level model. To be more precise, we use as service criterion the fraction of demand delivered directly from shelf, i.e. the P_2 service level. In this approach we assume that customer orders which can not be satisfied directly from shelf are backordered.

Dunsmuir and Snyder (1989) developed a simple model where intermittent demand is modelled as a compound Bernoulli process, that is, with a fixed probability there is positive demand during a time unit, otherwise demand is zero. However, they do not take into account the undershoot.

Basically, we will adapt the method presented by Dunsmuir and Snyder, such that the compound Bernoulli modelling is applicable for a more general class of situations. It is assumed that information is available on a daily basis. The review period, however, can be a week or a month, hence we also consider situations with a review period R which is larger than one. When the reorder quantity, Q, is large relative to the expected demand per time unit, or when the target service, $P_{2,target}$, is low, the target service can be realized, even when the reorder level, s, is negative. Therefore, we do not require the reorder point to be positive as in Dunsmuir and Snyder. In cases where Q and $P_{2,target}$ are both small, the shortages at the beginning of a replenishment cycle are relevant. Thus we do not neglect the expected shortage at the beginning of a replenishment cycle, contrary to Dunsmuir and Snyder. Finally, we also include the undershoot. The reason for this is that for demand processes which are not unit size, the undershoot has a considerable impact on performance levels, especially when the probability that demand is zero during the lead time is high. This is the case when demand is lumpy. However, when the demand is unit size the undershoot can be neglected. This extension of not neglecting the undershoot does not only improve the method presented by Dunsmuir and Snyder significantly, but also leads to a more complex expression for the service level.

The increasing importance of intermittent demand modelling can be argued as follows. We observed intermittent demand in inventory management of medicines in a medical centre with many departments. As management wants to keep inventories of certain medicines at a low echelon level (nursing departments), it had as a consequence that demand processes were intermittent on that level. Demand processes modelled as a compound Bernoulli process also appear in forecasting. When intermittent demand is forecasted, it appears (see e.g. Willemain et al. (1994)) that the separation procedure (called 'Croston's forecasting procedure') is better than the single exponential smoothing procedure, applied to the non-separated demand data. These conclusions also hold in all kind of data scenarios: interarrivals and demand occurrences cross-correlated or not, or interarrivals autocorrelated.

The compound Bernoulli assumption, with parameter π, was tested on a set of empirical data, obtained from a Dutch wholesaler of fasteners. In fact we tested whether the sample variance of the interarrival times of customers, $\hat{\sigma}_A^2$, is approximately equal to $\frac{1-\hat{\pi}}{\hat{\pi}^2}$, the estimated variance of the Bernoulli process. As estimate for π, $\hat{\pi}$, we used the reciprocal of the sample expected interarrival time. This was done for three classes of slow movers:

- C-class: one or more customer orders in 2 weeks
- D-class: one or more customer orders in one month
- E-class: one or more customer orders in one quarter

For each class a regression analysis, without intercept in the model, was performed, which resulted in

Class-C	$\frac{1-\hat{\pi}}{\hat{\pi}^2} = 0.934\hat{\sigma}_A^2$	adjusted $R^2 = 0.97$
Class-D	$\frac{1-\hat{\pi}}{\hat{\pi}^2} = 0.941\hat{\sigma}_A^2$	adjusted $R^2 = 0.95$
Class-E	$\frac{1-\hat{\pi}}{\hat{\pi}^2} = 0.966\hat{\sigma}_A^2$	adjusted $R^2 = 0.96$

Those results indicate that the compound Bernoulli assumption indeed might be a reasonable approximation for many intermittent demand processes.

This chapter is organized as follows. In section 3.1 a formal model description is given when the demand process is modelled as a compound Bernoulli process. In section 3.2 a method is presented for computing the reorder level s and the expected average physical stock for the same demand process. This method is called the compound Bernoulli method (CBM). In section 3.3 the CBM is validated by simulation, and the results will be compared with the results of the method presented by Dunsmuir and Snyder. Furthermore, examples are given how these result can be used by the management in practical situations. Finally, in section 3.4 some conclusions are given.

3.1 Model description

Let us assume that daily demand information is available. The demand size of the n-th day is denoted by D_n, and D_k^* denotes the demand size of the k-th day in which demand is positive. It is assumed that the D_n's as well as the D_k^*'s are independent and identically distributed random variables, with distribution functions $F_D(.)$ and $F_{D^*}(.)$, respectively.

When the demand process is modelled as a compound Bernoulli process and the probability that demand is positive is denoted by π_D, then the distribution functions $F_D(.)$ and $F_{D^*}(.)$ are related through

$$F_D(y) = \begin{cases} 1-\pi_D & \text{if} \quad y=0; \\ 1-\pi_D+\pi_D F_{D^*}(y) & \text{if} \quad y>0. \end{cases} \tag{3.1}$$

The relation between the moments of D and D^* can easily be derived:

$$\mathbb{E}D^k = \pi_D \mathbb{E}D^{*k} \quad k=1,2,\dots. \tag{3.2}$$

For small time units the compound Bernoulli process is approximately a compound Poisson process (see, for example, Feller (1970) page 153). Hence the compound Bernoulli process can be seen as the discrete time variant of the compound Poisson process.

We assume that the lead times do not cross in time, implying that the lead times of replenishment orders $L_1, L_2, \dots$ are dependent random variables (see, also, section 2.1.2). Furthermore, it is assumed that customer orders are handled during a day, whereas replenishment orders are handled at the end of the day.

Customer orders which cannot be delivered directly from stock will be backordered. As performance measure the P_2-service measure is used. Denote $S(R,s,Q)$ as the expected shortage during an arbitrary replenishment cycle, and $D(R,s,Q)$ as the expected demand during an arbitrary replenishment cycle, where a replenishment cycle is defined as the time between two successive arrivals of replenishment orders. Then, the following basic formula applies (see Tijms (1994) page 53).

$$P_2(R,s,Q) = 1 - \frac{S(R,s,Q)}{D(R,s,Q)}. \tag{3.3}$$

Another important performance measure is the expected average physical stock needed to maintain the required service level, $B_4(R,s,Q)$. Denote $H(R,s,Q)$ as the expected area between the physical stock level and the time-axis during a replenishment cycle, and $T(R,s,Q)$ as the expected duration of a replenishment cycle. Using the renewal reward theorem (see Tijms (1994) page 33) it can be seen that

$$B_4(R,s,Q) = \frac{H(R,s,Q)}{T(R,s,Q)}. \tag{3.4}$$

In order to derive expressions for $P_2(R,s,Q)$ and $B_4(R,s,Q)$ we define (see Figure 3.1)

$D(n)$	:=	the total demand during n subsequent time periods;
$D^*(n)$	:=	the total demand during n subsequent time periods, given that in at least one period the demand is positive;
T_k	:=	the point in time at which the inventory position drops below s for the k-th time after 0;
U_k	:=	s minus the inventory position at T_k (the k-th undershoot);
τ_k	:=	the first review moment after T_k;
W_k	:=	$\tau_k - T_k$;
$\hat{L}_k$	:=	$L_k + W_k$;
$U_{R,k}$	:=	s minus the inventory position at τ_k;
Z_k	:=	$D(\hat{L}_k) + U_k = D(L_k) + U_{R,k}$;
Z_k^*	:=	$D^*(\hat{L}_k) + U_k$.

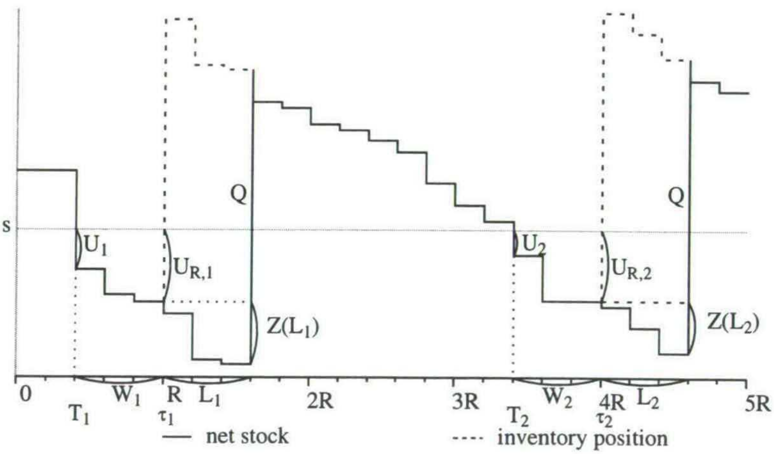


Figure 3.1: Evolution of the net stock and inventory position during the first replenishment cycle after zero.

Let zero be an arbitrary point in time, and consider the first replenishment cycle after zero, which can be considered as an arbitrary replenishment cycle. Analoguously to 2.27 we can derive that

$$P_2(R, s, Q) = 1 - \frac{\mathbb{E}(Z_2 - s)^+ - \mathbb{E}(Z_1 - s - Q)^+}{Q}. \tag{3.5}$$

To obtain an expression for $B_4(R, s, Q)$ we again consider the first replenishment cycle after zero. The following theorem gives an expression for $B_4(R, s, Q)$, which differs from (2.33).

Theorem 3.1 $$B_4(R,s,Q) \approx \frac{I\!E((s+Q-D(\hat{L}_1))^+)^2 - I\!E((s-D(\hat{L}_2))^+)^2}{2Q}.$$

Proof
Let $M(.)$ be the renewal function generated by the random process D_n, $n=1,2,\dots$ (with $0 < I\!ED < \infty$), and let $U(.)$ be the c.d.f. the asymptotic forwards recurrence time of D. Define $H(x)$ as the expected area between the physical inventory level and the zero level, in case the physical stock level on epoch 0 equals x ($x \geq 0$), and there are no replenishments. Then conditioning on the demand in the next period and using relation (3.1), results in

$$H(x) = x + \int_0^x H(x-y)dF_D(y). \tag{3.6}$$

Repeated substitution yields

$$H(x) = \int_0^x (x-y)dM(y). \tag{3.7}$$

In the sequel of the proof we will assume that $s > 0$, as for $s < 0$ the same approach can be applied. The expected physical stock at the beginning of the replenishment cycle (just after the replenishment arrived) at epoch $\tau_1 + L_1$, denoted by I_1, is equal to $s + Q - U_1 + D(\hat{L}_1)$. The expected physical stock at the end of the replenishment cycle (just before the replenishment arrives), denoted by I_2, is equal to $s - U_2 - D(\hat{L}_2)$. Then it is easy to see that $H(R,s,Q)$ is equal to $I\!EH(I_1) - I\!EH(I_2)$. Conditioning on I_1 and I_2, using (3.7) and Lemma 2.2, we find

$$\begin{aligned}
H(R,s,Q) &= \int_0^{s+Q} H(s+Q-x)d(F_{D(\hat{L}_1)} * U)(x) - \int_0^{s} H(s-x)d(F_{D(\hat{L}_2)} * U)(x) \\
&= \int_0^{s+Q}\int_0^{s+Q-x} (s+Q-x-y)dM(y)d(F_{D(\hat{L}_1)} * U)(x) \\
&\quad - \int_0^{s}\int_0^{s-x} (s-x-y)dM(y)d(F_{D(\hat{L}_2)} * U)(x) \\
&= \int_0^{s+Q}\int_0^{s+Q-x} (s+Q-x-y)d(M * U)(y)dF_{D(\hat{L}_1)}(x) \\
&\quad - \int_0^{s}\int_0^{s-x} (s-x-y)d(M * U)(y)dF_{D(\hat{L}_2)}(x) \\
&\approx \int_0^{s+Q}\int_0^{s+Q-x} \frac{s+Q-x-y}{I\!ED}dydF_{D(\hat{L}_1)}(x)
\end{aligned}$$

$$-\int_0^s \int_0^{s-x} \frac{s-x-y}{\mathbb{E}D} dy dF_{D(\hat{L}_2)}(x)$$
$$= \int_0^{s+Q} \frac{(s+Q-x)^2}{2\mathbb{E}D} dF_{D(\hat{L}_1)}(x) - \int_0^s \frac{(s-x)^2}{2\mathbb{E}D} dF_{D(\hat{L}_2)}(x).$$

Note that the expected duration of a replenishment cycle is given by $\frac{Q}{\mathbb{E}D}$, hence

$$\begin{aligned} B_4(R,s,Q) &\approx \int_0^{s+Q} \frac{(s+Q-x)^2}{2Q} dF_{D(\hat{L}_1)}(x) - \int_0^s \frac{(s-x)^2}{2Q} dF_{D(\hat{L}_2)}(x) \\ &= \frac{\mathbb{E}((s+Q-D(\hat{L}_1))^+)^2 - \mathbb{E}((s-D(\hat{L}_2))^+)^2}{2Q}. \end{aligned}$$

□

In the numerical experiments it is assumed that Z_1,Z_2, $D(\hat{L}_1)$, and $D(\hat{L}_2)$ are ME-distributed (see section 2.4). Hence the incomplete moments can be computed efficiently without using numerical integration. For given values for R, Q, and the target service level $P_{2,target}$, $P_2(R,s,Q) = P_{2,target}$ can be solved by using a local search procedure such as Golden Section search (see, for example, Press et al. (1992)). In case the demand process is a compound Bernoulli process there is a positive probability that $D(\hat{L}_1)$ (which is also a component of Z_1) is zero, therefore, it is doubtful to use a continuous distribution function as an approximation for the distribution function of $D(\hat{L}_1)$. In order to avoid this problem we explicitly make a distinction between the situation that $D(\hat{L}_1)$ is zero and positive.

3.2 The compound Bernoulli method

As has been argued in the previous section, the motivation behind the compound Bernoulli method is the distinction between the situations that the demand during $\hat{L}_k$ is zero or positive. Due to this distinction the expressions for $P_2(R,s,Q)$ and $B_4(R,s,Q)$ have to be adjusted. We denote the $\hat{L}_k$'s as the pseudo lead times and the probability of positive demand during the pseudo lead time as $\pi_{\hat{L}}$, i.e., $\pi_{\hat{L}} = \mathbb{P}(D(\hat{L}_k) > 0)$ for $k \geq 1$. In the situation that the demand during the pseudo lead time is zero, which occurs with probability $1-\pi_{\hat{L}}$, backlogs only occurs when the undershoot is larger than s, the value of the reorder point. However, for the situation that the demand during the pseudo lead time is positive, backlog occurs when the demand during the pseudo lead time (given it is positive) plus the undershoot is larger than s. Combining both possible situations, analogously to (3.5) the following relation can be derived, considering the first replenishment cycle after zero:

$$\begin{aligned}P_2(R,s,Q) &= 1-\pi_{\hat{L}}\frac{I\!E(Z_2^*-s)^+ - I\!E(Z_1^*-s-Q)^+}{Q}\\ &\quad -(1-\pi_{\hat{L}})\frac{I\!E(U_2-s)^+ - I\!E(U_1-s-Q)^+}{Q}.\end{aligned} \tag{3.8}$$

Analogously it can be derived that

$$\begin{aligned}B_4(R,s,Q) &\approx \pi_{\hat{L}}\frac{I\!E((s+Q-D^*(\hat{L}_1))^+)^2 - I\!E((s-D^*(\hat{L}_2))^+)^2}{2Q}\\ &\quad +(1-\pi_{\hat{L}})\frac{(s+Q)^2-s^2}{2Q}.\end{aligned} \tag{3.9}$$

To compute (3.8) and (3.9) we now may approximate the distribution functions of $D^*(\hat{L}_1)$, Z_1^* and U_1 by that of a mixture of two Erlang distribution. Hence, the first two moments of $D^*(\hat{L}_1)$, Z_1^*, and U_1, and the probability $\pi_{\hat{L}}$ are sufficient to calculate the $P_2(R,s,Q)$ and $B_4(R,s,Q)$ when R, s, and Q, are given.

For the first two moments of U_1 we use the asymptotic results for the first two moments of the forward recurrence time distribution, which yields according to (3.8)

$$I\!EU_1 \approx \frac{I\!E(D^*)^2}{2I\!ED^*} = \frac{I\!ED^2}{2I\!ED}, \tag{3.10}$$

$$I\!EU_1^2 \approx \frac{I\!E(D^*)^3}{3I\!ED^*} = \frac{I\!ED^3}{3I\!ED}. \tag{3.11}$$

Note that in order to compute the first two moments of U_1 we need the first three moments of D. Moreover, it is known from numerical investigations (see Tijms (1994)) that for practical purposes the relations (3.10) and (3.11) hold when (see 2.12)

$$Q \geq \begin{cases} \frac{3}{2}c_D^2 I\!ED & \text{if } c_D^2 > 1; \\ I\!ED & \text{if } 0.2 < c_D^2 \leq 1; \\ \frac{1}{2c_D}I\!ED & \text{if } 0 < c_D^2 \leq 0.2. \end{cases} \tag{3.12}$$

We calculate $\pi_{\hat{L}}$ from the generating function of $\hat{L}_1$, denoted by $P_{\hat{L}_1}(.)$, via

$$\begin{aligned}\pi_{\hat{L}} &= 1-I\!E(1-\pi_D)^{\hat{L}_1}\\ &= 1-P_{\hat{L}_1}(1-\pi_D).\end{aligned} \tag{3.13}$$

Because $\hat{L}_1$ is a convolution of two discrete random variables, a closed form expression for $P_{\hat{L}_1}(.)$ will not be available in general. Therefore we use a method described in Adan et al. (1995), to fit a discrete distribution function on a positive random variable using the first two moments of that random variable. The method of Adan et al. (1995) is summarize below. This method results in a distribution from one of the classes: Poisson, mixture of binomial, mixture of negative-binomial or mixture of geometric distributions. For any of these distributions the generating function can be determined. They also give bounds for the applicability of this discrete fit.

Lemma 3.2 *For a pair of non-negative, real numbers (μ_X, c_X). there exists a random variable X on the non-negative integers with mean μ_X and coefficient of variation c_X if and only if*

$$c_X^2 \geq \frac{2k+1}{\mu_x} - \frac{k(k+1)}{(\mu_X)^2} - 1 \tag{3.14}$$

where k is the unique integer satisfying $k \leq \mu_X < k+1$.

For the proof we refer to Adan et al. (1995). Let X a random variable on the non-negative integers, with mean $\mathbb{E}X$ and coefficent of variation c_X. Define $a := c_X^2 - 1/\mathbb{E}X$, then it follows from Lemma 3.2 that $a \geq -1$. The method is based on a selection out of four classes of distributions: Poisson, mixture of binomials, mixture of negative-binomials, and a mixture of geometric distributions. Define for $n = 0, 1, \ldots$

Poisson distribution	$P(\mu, n) := \sum_{i=0}^{n} \frac{\mu^i}{i!} e^{-\mu}$
Negative-binomial distribution	$NB(k, p, n) := \sum_{i=0}^{n} \binom{k+i-1}{k-1} p^k (1-p)^i$
Binomial distribution	$BIN(k, p, m) := \sum_{i=0}^{m} \binom{k}{i} p^i (1-p)^{k-i} \quad m = 0, 1, \ldots, k$
Geometric distribution	$G(p, n) := \sum_{i=0}^{n} p(1-p)^i$

Then there exists a random variable Y which matches the first two moments of X, if the distribution function of Y is chosen such that:

- If $\frac{-1}{k} \leq a \leq \frac{-1}{k+1}$ then

  $$\mathbb{P}(Y \leq n) = qBIN(k, p, n) + (1-q)BIN(k+1, p, n) \quad (n = 0, 1, \ldots, k+1),$$

  where

  $$\begin{aligned} q &= \frac{1 + a(1+k) + \sqrt{-ak(1+k) - k}}{1+a} \\ p &= 1 - \frac{\mathbb{E}X}{k+1-q}. \end{aligned}$$

- If $a = 0$ then

  $$\mathbb{P}(Y \leq n) = P(\mu, n) \quad (n = 0, 1, \ldots),$$

  where $\mu = 1/\mathbb{E}X$.

- If $\frac{1}{k+1} \leq a \leq \frac{1}{k}$ then

  $$\mathbb{P}(Y \leq n) = qNB(k, p, n) + (1-q)NB(k+1, p, n) \quad (n = 0, 1, \ldots, k+1),$$

where

$$\begin{aligned} q &= \frac{a(1+k)-\sqrt{(1+k)(1-ak)}}{1+a}, \\ p &= 1-\frac{\mathbb{E}X}{k+1-q+\mathbb{E}X}. \end{aligned}$$

- If $a \geq 1$ then

$$\mathbb{P}(Y \leq n) = qG(p_1, n) + (1-q)G(p_2, n) \quad (n = 0, 1, \ldots),$$

where

$$\begin{aligned} q &= \frac{1}{1+a+\sqrt{a^2-1}}, \\ p_1 &= \frac{2}{2+\mathbb{E}X(1+a+\sqrt{a^2-1})}, \\ p_2 &= \frac{2}{2+\mathbb{E}X(1+a-\sqrt{a^2-1})}. \end{aligned}$$

To obtain the first two moments of $\hat{L}_1$ we need the first two moments of L_1, which are assumed to be given, and the first two moments of W_1.

Theorem 3.3 *W_1 is uniformly distributed over $\{0, 1, \ldots, R-1\}$.*

Proof
Define $W(x)$ as the time between the moment of undershoot of the reorder level s for the first time after zero and the next review epoch, given that the inventory position equals $s+x$ at time epoch 0, where $x > 0$.
Then for $0 \leq k < R$

$$\begin{aligned} \mathbb{P}(W(x)=k) &= \sum_{m=1}^{\infty} \mathbb{P}(\sum_{n=1}^{mR-k} D_n > x, \sum_{n=1}^{mR-k-1} D_n \leq x) \\ &= \sum_{m=1}^{\infty} \Big(\mathbb{P}(\sum_{n=1}^{mR-k-1} D_n \leq x) - \mathbb{P}(\sum_{n=1}^{mR-k-1} D_n \leq x, \sum_{n=1}^{mR-k} D_n \leq x)\Big) \\ &= \sum_{m=1}^{\infty} \Big(F_{D_1}^{(mR-k-1)*}(x) - F_{D_1}^{(mR-k)*}(x)\Big). \end{aligned} \tag{3.15}$$

Taking the Laplace transforms at both sides yields

$$\tilde{W}_k(s) := \int_0^{\infty} e^{-sx} \mathbb{P}(W(x)=k)dx$$

$$
\begin{aligned}
&= \frac{1}{s}\int_0^\infty e^{-sx} d_x \mathbb{P}(W(x)=k) \\
&= \frac{1}{s}\sum_{m=1}^{\infty}\Big(\int_0^\infty e^{-sx} dF_{D_1}^{(mR-k-1)*}(x) - \int_0^\infty e^{-sx} dF_{D_1}^{(mR-k)*}(x)\Big) \\
&= \tilde{F}_{D_1}^{R-k}(s)\frac{1-\tilde{F}_{D_1}(s)}{s(1-\tilde{F}_{D_1}^{R}(s))},
\end{aligned}
\tag{3.16}
$$

where $\tilde{F}_{D_1}(s) := \int_0^\infty e^{-sx} dF_{D_1}(x)$.
Since

$$\lim_{x\to\infty} \mathbb{P}(W(x)=k) = \lim_{s\downarrow 0} s\tilde{W}_k(s), \tag{3.17}$$

we conclude from (3.16) that

$$\lim_{x\to\infty} \mathbb{P}(W(x)=k) = \lim_{s\downarrow 0} \tilde{F}_{D_1}^{R-k}(s)\frac{1-\tilde{F}_{D_1}(s)}{1-\tilde{F}_{D_1}^{R}(s)}. \tag{3.18}$$

This implies, using l'Hopital's rule

$$\lim_{x\to\infty} \mathbb{P}(W(x)=k) = \frac{1}{R}, \tag{3.19}$$

which completes the proof. □

Hence

$$
\begin{aligned}
\mathbb{E}W_1 &= \tfrac{1}{2}(R-1), && (3.20)\\
\mathbb{E}W_1^2 &= \tfrac{1}{6}(R-1)(2R-1). && (3.21)
\end{aligned}
$$

Because L_1 and W_1 are independent, the following basic formulas apply

$$
\begin{aligned}
\mathbb{E}\hat{L}_1 &= \mathbb{E}L_1 + \mathbb{E}W_1, && (3.22)\\
\mathbb{E}\hat{L}_1^2 &= \mathbb{E}L_1^2 + 2\mathbb{E}L_1\mathbb{E}W_1 + \mathbb{E}W_1^2. && (3.23)
\end{aligned}
$$

Since $D(\hat{L}_1)$ is a stochastic sum of i.i.d. random variables, we have

$$
\begin{aligned}
\mathbb{E}D(\hat{L}_1) &= \mathbb{E}\hat{L}_1\mathbb{E}D && (3.24)\\
\sigma^2(D(\hat{L}_1)) &= \mathbb{E}\hat{L}_1\sigma^2(D) + \sigma^2(\hat{L}_1)(\mathbb{E}D)^2 && (3.25)
\end{aligned}
$$

What remains are expressions for the first two moments of Z_1^*, and therefore expressions for the first two moments of $Z^*(\hat{L}_1)$. Using the definition of $\pi_{\hat{L}}$ and the appropriate analogy of relation (3.1) it can be shown that

$$
\begin{aligned}
\mathbb{E}D^*(\hat{L}_1) &= \frac{\mathbb{E}D(\hat{L}_1)}{\pi_{\hat{L}}}, && (3.26)\\
\sigma^2(D^*(\hat{L}_1)) &= \frac{\sigma^2(D(\hat{L}))}{\pi_{\hat{L}}} - \frac{(1-\pi_{\hat{L}})\mathbb{E}D(\hat{L}_1)^2}{\pi_{\hat{L}}^2}. && (3.27)
\end{aligned}
$$

From $\sigma^2(D^*(\hat{L}_1)) \geq 0$ it follows, by using (3.27), that $c^2_{D(\hat{L_1})} \geq (1-\pi_{\hat{L}})/\pi_{\hat{L}}$. Hence, when $c^2_{D(\hat{L_1})} < (1-\pi_{\hat{L}})/\pi_{\hat{L}}$, the compound Bernoulli model can not be applied. In this situation we propose to use expressions (3.10),(3.11),(3.24) and (3.25) to obtain values for the first two moments of Z_1, and use service equation (3.5) to calculate the reorder point s.

3.3 Numerical results

To show the impact of the extensions of the CBM with respect to the method presented by Dunsmuir and Snyder, we use all cases that are considered in their paper. The reorder levels calculated according to Dunsmuir and Snyder as well as the reorder level calculated by the CBM are both validated by simulation. Putting it more precisely, the actual resulting service level is computed via simulation, given a value for the reorder point, and this level is compared to the required service level. The closer these two levels are to each other the better the method performs. The number of sub-runs is fixed at 10 (exclusive the initialisation run), and the sub-run length is 100.000 time units. Furthermore, the demand sizes of the customer and lead times are drawn out of mixtures of Erlang distributions. But since we assumed that replenishment orders did not cross we have to be careful with the generation of lead times, see section 2.1.2. The results are tabulated in Table 3.1, in which s_1 denotes the reorder point calculated by Dunsmuir and Snyder with the associated actual service level β_1 (between brackets the 95% confidence interval is given), and s_2 denotes the reorder point calculated by the CBM with the associated actual service level β_2.
The cases considered by Dunsmuir and Snyder are such that the undershoot has a consider-

Table 3.1: Simulation results for $P_{2,target} = 0.95$, $(\mathbb{E}L_1\,;\,\sigma(L_1)) = (2\,;\,0)$ and $R = 1$

				Dunsmuir & Snyder		CBM	
π_D	$\mathbb{E}D^*$	$\sigma(D^*)$	Q	s_1	β_1	s_2	β_2
0.36	3.00	1.41	2	6.00	0.8521 (± 0.0010)	8.14	0.9480 (± 0.0011)
			3	5.00	0.8115 (± 0.0010)	7.74	0.9481 (± 0.0008)
			4	4.30	0.7891 (± 0.0016)	7.38	0.9485 (± 0.0013)
0.28	10.30	3.51	5	18.30	0.8586 (± 0.0022)	24.15	0.9477 (± 0.0016)
			7	17.00	0.8492 (± 0.0023)	23.32	0.9479 (± 0.0015)
			10	15.00	0.8324 (± 0.0026)	22.17	0.9480 (± 0.0015)
0.45	201.60	212.40	200	730.00	0.8956 (± 0.0037)	942.24	0.9492 (± 0.0031)
			300	640.00	0.8782 (± 0.0038)	898.73	0.9490 (± 0.0028)
			400	570.00	0.8670 (± 0.0037)	858.56	0.9493 (± 0.0029)
0.64	846.60	384.60	1100	1975.00	0.8517 (± 0.0022)	2575.06	0.9509 (± 0.0008)
			1700	1725.00	0.8444 (± 0.0018)	2384.73	0.9502 (± 0.0008)
			2200	1600.00	0.8519 (± 0.0018)	2251.34	0.9499 (± 0.0009)

able impact on the service levels, which is shown by the bad service performance in case the undershoot is neglected ($\beta_1 < P_{2,target} = 0.95$). In contrast with this, the CBM has an excellent performance for all cases considered by Dunsmuir and Snyder ($\beta_2 \approx P_{2,target} = 0.95$), in spite of the small values of Q when compared with $\mathbb{E}D^*$.

Next we use simulation to validate the quality of the CBM in terms of service performance and expected average physical stock for a wide range of parameter values. The results are given in Table 3.2 and 3.3, in which s_1 denotes the reorder point calculated by the CBM with the associated actual service level β_1 (between brackets the 95% confidence interval is given), and $B_4(R, s, Q)$ the expected physical stock calculated by expression (3.9) with the associated actual physical stock μ_1. Furthermore, we compared μ_1 also with an often used simple expression for the expected average physical stock, namely $\mu_2 = s - \mathbb{E}Z_1 + Q/2$. From these simulation results it can be concluded that the performance of the CBM is excellent for most situations considered, whereas the simple expression μ_2 deviates from μ_1 in all cases. However, for the situation that $\pi_D = 0.9$ and $(\mathbb{E}L_1\,;\,\sigma(L_1)) = (10\,;\,4)$ (see bold printed results in Table 3.2) the actual service is too large. A possible explanation for this is that in these situations the value of Q is too small with respect to the value of U_R. This violates the previous made assumption that never a multiple of Q is replenished. Moreover, the expressions for the undershoot ((3.10) and (3.11)), as well as the result that W_1 is distributed uniformly on $\{0, 1, .., R-1\}$, are based on asymptotic results from renewal theory which only hold for sufficiently large values of Q.

Table 3.2: Results to validate the CBM ($\mathbb{E}D^* = 5$, $\sigma(D^*) = 5$, and $\mathbb{E}L_1 = 10$)

R	π_D	Q	$(\sigma(L),\sigma(\hat{L}))$	$P_{2,target}$	β_1	s_1	B_4	μ_1	μ_2
5	0.10	50	(4.00 ; 4.03)	0.95	0.9497 (± 0.0013)	14.11	33.27	33.21	29.11
5	0.10	50	(4.00 ; 4.00)	0.99	0.9898 (± 0.0006)	27.31	46.33	46.31	42.31
5	0.90	50	(4.00 ; 4.01)	0.95	**0.9587** (± 0.0009)	87.42	59.07	58.67	62.42
5	0.90	50	(4.00 ; 4.02)	0.99	**0.9947** (± 0.0003)	116.61	87.72	87.29	91.61
5	0.10	200	(4.00 ; 3.94)	0.95	0.9510 (± 0.0013)	1.03	95.25	95.03	91.03
5	0.10	200	(4.00 ; 3.92)	0.99	0.9900 (± 0.0008)	16.03	110.06	109.78	106.03
5	0.90	200	(4.00 ; 4.04)	0.95	0.9490 (± 0.0007)	59.65	106.42	106.05	109.65
5	0.90	200	(4.00 ; 4.01)	0.99	0.9907 (± 0.0004)	92.73	138.86	138.52	142.73
10	0.10	50	(4.00 ; 4.02)	0.95	0.9482 (± 0.0019)	16.25	34.18	34.07	31.25
10	0.10	50	(4.00 ; 4.01)	0.99	0.9902 (± 0.0008)	30.14	47.92	47.91	45.14
10	0.90	50	(4.00 ; 4.01)	0.95	**0.9548** (± 0.0008)	104.17	64.69	64.31	79.17
10	0.90	50	(4.00 ; 4.02)	0.99	**0.9936** (± 0.0002)	136.29	96.17	95.83	111.28
10	0.10	200	(4.00 ; 4.06)	0.95	0.9497 (± 0.0009)	2.40	95.39	95.48	92.40
10	0.10	200	(4.00 ; 4.07)	0.99	0.9891 (± 0.0006)	18.28	111.06	110.98	108.28
10	0.90	200	(4.00 ; 4.00)	0.95	0.9488 (± 0.0008)	74.12	109.75	109.20	124.12
10	0.90	200	(4.00 ; 3.99)	0.99	0.9909 (± 0.0002)	110.34	145.24	144.99	160.34

Table 3.3: Results to validate the CBM ($\mathbb{E}D^* = 5$ and $(\mathbb{E}L_1\,;\,\sigma(L_1)) = (1{,}0)$)

R	π_D	$\sigma(D^*)$	Q	$P_{2,target}$	β_1	s_1	B_4	μ_1	μ_2
1	0.10	5	10	0.99	0.9902 (± 0.0015)	20.81	25.31	25.30	20.31
1	0.90	5	10	0.99	0.9899 (± 0.0009)	28.37	28.87	28.86	23.87
1	0.10	10	10	0.99	0.9914 (± 0.0022)	65.60	70.10	70.11	57.60
1	0.90	10	10	0.99	0.9901 (± 0.0014)	76.44	76.95	76.94	64.44
5	0.10	5	10	0.95	0.9501 (± 0.0037)	14.75	18.32	18.30	14.25
5	0.90	5	10	0.95	0.9515 (± 0.0013)	36.53	28.24	28.18	32.03
5	0.10	10	10	0.95	0.9520 (± 0.0047)	41.66	45.23	45.21	33.66
5	0.90	10	10	0.95	0.9492 (± 0.0026)	66.99	58.85	58.76	54.99
1	0.10	5	50	0.95	0.9500 (± 0.0023)	4.32	28.84	28.77	23.82
1	0.90	5	50	0.95	0.9497 (± 0.0012)	10.01	30.57	30.57	25.51
1	0.10	10	50	0.95	0.9521 (± 0.0042)	24.84	49.36	49.42	36.84
1	0.90	10	50	0.95	0.9489 (± 0.0026)	32.83	53.45	53.39	40.83
5	0.10	5	50	0.99	0.9901 (± 0.0009)	16.03	39.54	39.49	35.53
5	0.90	5	50	0.99	0.9915 (± 0.0010)	40.20	51.74	51.71	55.70
5	0.10	10	50	0.99	0.9911 (± 0.0023)	54.68	78.19	78.18	66.68
5	0.90	10	50	0.99	0.9898 (± 0.0017)	84.72	96.29	96.16	92.72
10	0.10	5	200	0.95	0.9489 (± 0.0011)	-2.25	95.13	94.89	92.25
10	0.90	5	200	0.95	0.9472 (± 0.0007)	23.73	99.52	99.40	114.23
10	0.10	10	200	0.95	0.9495 (± 0.0020)	6.85	104.23	103.66	93.85
10	0.90	10	200	0.95	0.9491 (± 0.0007)	41.00	116.93	116.91	124.00

To illustrate that negative values for the reorder level may be appropriate, we consider two situations. In the first situation a low service is required, whereas in the second situation the reorder quantity is large relative to $\mathbb{E}D^*$, see Table 3.4 and 3.5. Moreover, notice the excellent results of the CBM in these situations of both the reorder level and the expected average physical stock, wheras μ_2 even is negative!

Table 3.4: Results to illustrate negative reorder levels $R = 1$, $Q = 50$, $\mathbb{E}D^* = 5$, and $P_{2,target} = 0.50$

π_D	$\sigma(D^*)$	$(\mathbb{E}L_1;\sigma(L_1))$	β_1	s_1	B_4	μ_1	μ_2
0.10	5	(1 ; 0)	0.5022 ($\pm$ 0.0045)	-19.51	9.04	9.06	-0.01
0.10	5	(10 ; 4)	0.5028 ($\pm$ 0.0048)	-15.13	9.43	9.47	-0.13
0.90	5	(1 ; 0)	0.5004 ($\pm$ 0.0016)	-15.54	9.22	9.22	-0.04
0.90	5	(10 ; 4)	0.4844 ($\pm$ 0.0041)	22.46	12.35	11.99	-2.54
0.10	10	(1 ; 0)	0.4983 ($\pm$ 0.0069)	-13.28	13.23	13.24	-1.28
0.10	10	(10 ; 4)	0.4995 ($\pm$ 0.0079)	-9.52	13.70	13.70	-2.02
0.90	10	(1 ; 0)	0.4989 ($\pm$ 0.0031)	-9.76	13.62	13.61	-1.76
0.90	10	(10 ; 4)	0.4845 ($\pm$ 0.0044)	25.81	18.14	17.43	-6.69

Table 3.5: Results to illustrate negative reorder levels $R = 1$, $Q = 500$, $\mathbb{E}D^* = 5$, and $P_{2,target} = 0.90$

π_D	$\sigma(D^*)$	$(\mathbb{E}L_1;\sigma(L_1))$	β_1	s_1	B_4	μ_1	μ_2
0.10	5	(1 ; 0)	0.9003 ($\pm$ 0.0017)	-44.49	207.04	206.18	200.01
0.10	5	(10 ; 4)	0.8995 ($\pm$ 0.0021)	-40.02	207.06	206.32	199.99
0.90	5	(1 ; 0)	0.8999 ($\pm$ 0.0004)	-40.50	207.05	206.81	200.00
0.90	5	(10 ; 4)	0.8996 ($\pm$ 0.0025)	0.00	207.60	207.27	200.00
0.10	10	(1 ; 0)	0.8994 ($\pm$ 0.0041)	-37.00	213.92	212.88	200.00
0.10	10	(10 ; 4)	0.8990 ($\pm$ 0.0048)	-32.51	214.03	212.98	200.00
0.90	10	(1 ; 0)	0.8995 ($\pm$ 0.0014)	-33.01	213.99	213.99	199.99
0.90	10	(10 ; 4)	0.8993 ($\pm$ 0.0025)	7.57	215.21	215.04	200.07

To conclude we indicate some restrictions to the application of the compound Bernoulli model. When the demand process is a compound renewal process, it is in general not true that the D_n's are independent identically distributed. Secondly, when the probability distribution of the demand size is concentrated in a small number of points, it is in general not correct to assume that the distribution function of Z_1, $D(\hat{L}_1)$ and U_1 are mixtures of two Erlang distributions. Moreover, the expressions for the undershoot ((3.10) and (3.11)) as well as the result that W_1 is distributed uniformly on $\{0, 1, .., R-1\}$, are based on asymptotic results from renewal theory. Hence, for values of Q small as compared to $\mathbb{E}D$, these relations do not hold.

From a managerial point of view it is interesting to represent graphically various service levels versus the expected average physical stock level, see Figures 3.2 and 3.3. In Figures 3.2 and 3.3 we consider the situations with $R = 1$, $Q = 50$, $\mathbb{E}D^* = 5$, $\pi_D = 0.5$, $\mathbb{E}L_1 = 5$, and $\sigma(L_1) = 2$. In order to make a trade off between the customer service and the associated required average physical stock, the graph can be used as an aid for determining the target service level. For the determination of the replenishment quantity Q often the economic order quantity is used (see, e.g., see Silver and Peterson (1985)), which is also known as the Wilson lot size formula. A more sophisticated approach would be to minimize the ordering plus holding costs subject to the service level constraint (see, for example, Moon and Choi

(1994)). The expected total relevant cost during a replenishment cycle can now be written as

$$C(R, s, Q) = \frac{a\mathbb{E}D}{Q} + b_4 B_4(R, s, Q) \tag{3.28}$$

The problem can now be formulated as

$$\begin{array}{lll} (\mathcal{P}_1) & \text{minimize} & C(R, s, Q) \\ & \text{s.t.} & P_2(R, s, Q) = P_{2,target}; \\ & & 0 \leq Q, s \in \mathbb{R}. \end{array}$$

where b_4 denotes the stock keeping costs per unit per unit of time, and a denotes the fixed ordering costs per replenishment. Note that for any given Q, the minimal value for s can be determined by solving $P_2(R, s, Q) = P_{2,target}$. Let $s^*(Q, P_{2,target})$ denote the optimal value of s as function of Q and $P_{2,target}$. Then $(\mathcal{P}_1)$ can be reformulated into the following one-dimensional optimization problem

$$\begin{array}{lll} (\mathcal{P}_2) & \text{minimize} & C(s^*(Q, P_{2,target}), Q) \\ & & 0 \leq Q. \end{array}$$

Provided that $C(s^*(Q, P_{2,target}), Q)$ is convex, $(\mathcal{P}_2)$ can be solved by using a local search procedure (see, for example Press et al. (1992)).

Consider the following example, $R = 1$, $(\mathbb{E}D^*; \sigma(D^*)) = (5; 5)$, $\pi_D = 0.50$, $(\mathbb{E}L; \sigma(L)) = (10; 2)$, $P_{2,target} = 0.95$, $a = 50$ \$ and b_4 varies between 1, 5 and 10 \$ / year (= 200 days). Note that in this case the EOQ is given by $\sqrt{10000/b_4}$. In Table 3.6 the results are given for the optimal replenishment quantity Q^* of the optimization problem described above. Figure 3.4 shows that the solution for Q^* is robust for small values of b_4. Hence, in these situations the EOQ is nearly optimal.

Table 3.6: The optimal replenishment quantities as function of the holding costs

b_4	Q_{EOQ}	$C(s^*(Q_{EOQ}), Q_{EOQ})$	Q^*	$C(s^*(Q^*), Q^*)$	$\Delta C(s^*(Q), Q)$
10	100	1078.5	114	1070.5	8.0
5	141	723.7	157	719.5	4.2
1	316	306.6	335	306.1	0.5

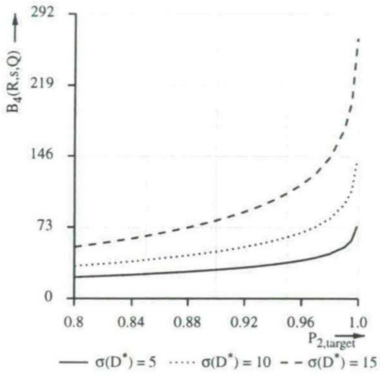


Figure 3.2: Average physical stock level for $P_{2,target} \in (0.80, ..., 0.99)$).

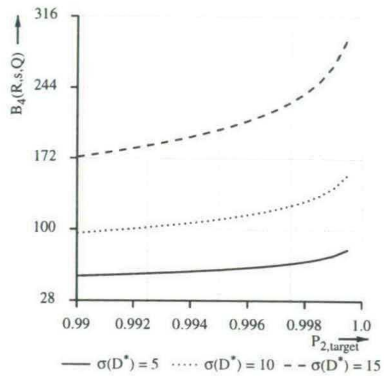


Figure 3.3: Average physical stock level for $P_{2,target} \in (0.99, ..., 0.999)$.

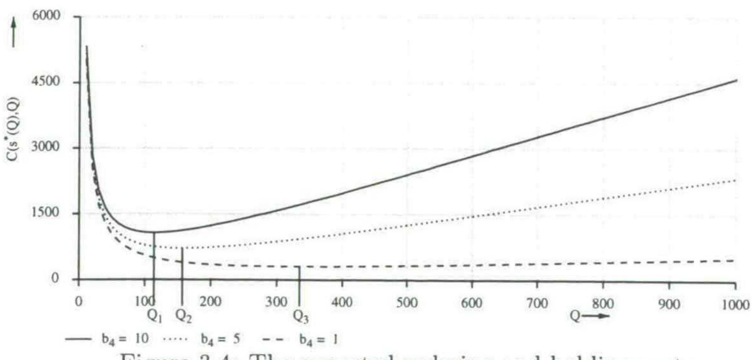


Figure 3.4: The expected ordering and holding costs.

3.4 Conclusions and future research

In this chapter we developed a method for the determination of the reorder point s and the average physical stock level in an (R, s, Q) inventory model subject to a service level constraint when the demand process is intermittent. Toward this end, we modelled the demand process as a compound Bernoulli process. The motivation behind the compound Bernoulli model is the distinction between the situations that the demand during the pseudo lead time is zero or positive. The method presented is an extension of the method introduced by Dunsmuir and Snyder (1989). The most important extension is the incorporation of the undershoot, which is neglected by Dunsmuir and Snyder. The undershoot has a considerable impact on the performance measures, as is shown by numerical experiments. The CBM turned out to perform excellent in almost all situations considered. The quality of the simple expression for the expected average physical stock ($\mu_2 = s - \mathbb{E}Z_1 + Q/2$) is poor, therefore we advise the use of the more complex, although computationally simple, formula (3.9). Finally, illustrative examples are given how the CBM can be used by management in practical situations.

Chapter 4

The impact of data collection on P_2 service level

This chapter is based on Janssen, Heuts and de Kok (1996). Although the concept of the (R, s, Q) inventory model is well-known, both in theory as well as in practice, little is known about the impact of the modelling assumptions with respect to the demand process on the customer service level.

Under the regime of the (R, s, Q) inventory policy the inventory position is monitored every R time units in order to consider the necessity of a replenishment decision. When the inventory position is below s, an amount of Q units is ordered such that the inventory position is raised to a value between s and $s + Q$ (see section 2.1.3).

The assumptions about the demand process are usually determining the complexity of the inventory model. The simplest assumption is that the demand is constant and known. Although these assumptions are quite restrictive, models requiring these assumptions are still important. First, many results are quite robust with respect to the model parameters, as is illustrated by the economic order quantity (EOQ). Secondly, the results of these simple models are often good starting solutions for the more complex models.

When demand is uncertain the demand process can be described by a discrete time model or a continuous time model. For both models we assume that the c.d.f. for the demand is known, and that there is a history of past observations from which the form and parameter values of the demand distribution are estimated. The discrete time models are extensively described in the literature (see, e.g., Schneider (1981, 1990), Tijms and Groenevelt (1984), Silver and Peterson (1985), and Tersine (1994)). The discrete time models assume that the time axis is divided in disjunct time units of length T, for example, days. Moreover, it is assumed that information is available about the first two moments of the demand per time unit (obtained from historical data). To reduce the complexity of the model it is assumed that the demands per time unit are independent and identically distributed random variables (in general this random variable might have positive probability mass at zero). A method which much resembles this model is the model described by Dunsmuir and Snyder (1989), where the demand is modelled as a compound Bernoulli process, i.e. with a fixed probability there is positive demand during a time unit, else demand is zero (see Chapter 3).

In the continuous time models the time axis is *not* divided in disjunct intervals. The demand process can be described as a compound Poisson process or a compound renewal process. For the (s, S) model where demand is modelled as a compound renewal process we refer to Sahin (1983, 1990). Continuous time review has also been referred to as transaction reporting (see, for example, Hadley and Whitin (1963)). For the continuous time models it is assumed that information is available about the first two moments of the interarrival times of customers as well as about the first two moments of the demand sizes of each customer. It is clear that the continuous time models requires specific and more detailed customer information. In principle this information is available, yet most business information systems do not collect demand data on a customer by customer basis. The discrete time models require only information about demands per time unit, which is usually available in MRP-systems.

In this chapter we investigate to what extent a continuous time demand process can be approximated by a discrete time process. For practical purposes the interesting question arises how the customer service level is influenced when the true demand process is described by a compound renewal process, whereas the demand is described by a discrete time model to determine the decision variables. Furthermore, we present a continuous time model to determine approximations for some performance measures. We validate the quality of these approximations by simulation.

The organisation of this chapter is as follows. In section 4.1 the general assumptions of the (R, s, Q) inventory model are presented. In sections 4.2 and 4.3 the continuous time model and the discrete time model are described, while in section 4.4 several numerical comparisons can be found based on discrete event simulation. Finally, section 4.5 summarizes our conclusions.

4.1 Model description

In order to specify the (R, s, Q) inventory model we distinguish between the demand process and the lead time process. We assume that the demand process is a compound renewal process. That is, the interarrival times of customers are described by the sequence $\{A_i\}_{i=1}^{\infty}$ of independent and identically distributed (i.i.d.) random variables with a common distribution function F_A, where A_i represents the time between the arrival of the i-th and $(i-1)$-th customer after time epoch 0. Furthermore, we assume that the process is stationary at time epoch 0, which is assumed to be a review moment. Time epoch 0 is considered to be an arbitrary point in time with respect to the renewal process associated with $\{A_i\}_{i=1}^{\infty}$, which implies that A_1 is distributed according to a residual lifetime distribution. The demand sizes of the customers are described by the sequence $\{D_i\}_{i=1}^{\infty}$ of i.i.d. random variables with a common distribution function F_D, where D_i represents the demand size of the i-th customer after time epoch 0. The sequence $\{D_i\}_{i=1}^{\infty}$ is independent of $\{A_i\}_{i=1}^{\infty}$.

Shortages are backordered, and it is assumed that per review period at most one time

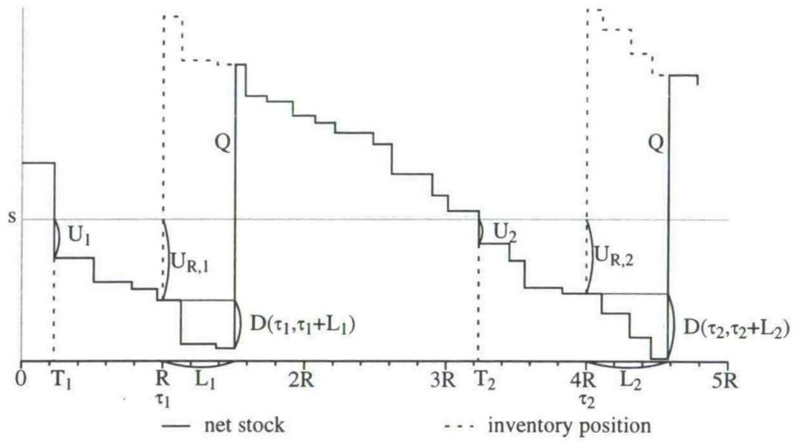


Figure 4.1: Evolution of the net stock and inventory position during the first replenishment cycle

the quantity Q is ordered. Furthermore, we assume that lead times of subsequent replenishment orders do not cross in time.

The problem discussed in this chapter considers the determination of the minimal value of the reorder level s, such that a target service level is achieved for given values for Q and R. We define the following variables, confirm the definitions in the previous chapters:

$N(t_1,t_2)$	$:=$	the number of customer arrivals in $(t_1,t_2]$;
$D(t_1,t_2)$	$:=$	the total demand in $(t_1,t_2]$;
$D_{t,i}$	$:=$	$D(it,(i+1)t)$ where $t \in I\!R^+$ and $i \in I\!N$;
$X(t)$	$:=$	the inventory position at time t;
T_k	$:=$	the point in time at which the inventory position drops below s the k-th time after 0;
U_k	$:=$	s minus the inventory position just after T_k (the k-th undershoot);
τ_k	$:=$	the first review moment after T_k;
$U_{R,k}$	$:=$	s minus the inventory position just before τ_k;
Z_k	$:=$	$D(\tau_k,\tau_k+L_k)+U_{R,k}$.

We now focus on an arbitrary replenishment cycle. Without loss of generality we can concentrate on the first complete replenishment cycle after zero (see, for example, Hadley and Whitin (1963) or de Kok (1991c)). Analogously to (2.27) we have that

$$P_2(R,s,Q) = 1 - \frac{I\!E(Z_2-s)^+ - I\!E(Z_1-s-Q)^+}{Q}. \tag{4.1}$$

Because all the processes involved are stationary, we conclude that $D(\tau_1,\tau_1+L_1) \stackrel{d}{=} D(\tau_2,\tau_2+L_2)$ and $U_{R,1} \stackrel{d}{=} U_{R,2}$, hence $Z_1 \stackrel{d}{=} Z_2$.

For given values for R, Q and the target service level (denoted by $P_{2,target}$), the optimal value of s can be determined by solving $P_2(R,s,Q) = P_{2,target}$. To avoid numerical integration when computing $\mathbb{E}(Z_1 - s)^+$ and $\mathbb{E}(Z_1 - s - Q)^+$, we approximate the distribution of Z_1 by a ME-distributions, see section 2.4. In the literature it is shown that the demand distribution during lead time or review period plus lead time, can well be approximated by a gamma distribution (see Burgin (1975)). Our own practical investigations confirm the applicability of this assumption. Hence, to calculate the reorder point s only the first two moments of Z_1 are required. Basically, the discrete time model and the continuous time model differ in the way these moments are obtained.

4.2 The continuous time model

As has been argued above, the first two moments of Z_1 are required in order to compute the optimal value of s. Since $D(\tau_1, \tau_1 + L_1)$ and $U_{R,1}$ are independent random variables, it is sufficient to derive expression for the moments of $D(\tau_1, \tau_1 + L_1)$ and $U_{R,1}$ separately. In this section we will work with the following assumptions, for which the reasons are explained in the sequel of the section

Condition 4.1

C.1 $Q \geq Cond(D_{R,0})$;

C.2 $R \geq Cond(A)$;

C.3 $c_A^2 \in [0, \frac{3R}{\mathbb{E}A} + \sqrt{\frac{9R^2}{(\mathbb{E}A)^2} + \frac{6R}{\mathbb{E}A}c_D^2 + 1})$;

C.4 $\mathbb{E}L \geq Cond(A)$;

C.5 $c_A^2 \in [0, \frac{3\mathbb{E}L}{\mathbb{E}A} + \sqrt{\frac{9(\mathbb{E}L)^2}{(\mathbb{E}A)^2} + \frac{6\sigma(L)^2}{(\mathbb{E}A)^2} + \frac{6\mathbb{E}L}{\mathbb{E}A}c_D^2 + 1})$,

where $Cond(.)$ *is given by (2.12).*

The distribution function of $U_{R,1}$ can be approximated by the asymptotic forward recurrence time distribution of the renewal process generated by the sequence of independent and identically distributed random variables with common distribution function $F_{D_{R,0}}$. Using Theorem 2.1 for $k = 1$ and $k = 2$ yields

$$\mathbb{E}U_{R,1} \approx \frac{\mathbb{E}D_{R,0}^2}{2\mathbb{E}D_{R,0}}, \tag{4.2}$$

$$\mathbb{E}U_{R,1}^2 \approx \frac{\mathbb{E}D_{R,0}^3}{3\mathbb{E}D_{R,0}}. \tag{4.3}$$

These relations are only valid when Condition $C.1$ ($Q \geq Cond(D_{R,0})$) holds.

The calculation of (4.2) and (4.3) requires the first three moments of $D_{R,0}$. In view of computing the third moment of $D_{R,0}$ there are two approaches that can be followed.

Namely, we can derive an expression for $\mathbb{E}D^3_{R,0}$, involving the first three moments of D and $N(0,R)$. For the other approach, we approximate the distribution function of $D_{R,0}$ by a gamma distribution. In that case the third moment can be calculated when the first two moments are known. The estimates of forecasts of higher moments (third and higher) of a stochastic variable are very sensitive for extreme values in the data. To avoid the use of third moments and higher of D and $N(0,R)$, we will follow the second approach. Hence, we only need the first two moments of $D_{R,0}$. Using (2.14) and (2.15) with t equal to R yields,

$$\begin{aligned} \mathbb{E}D_{R,0} &= \mathbb{E}N(0,R)\,\mathbb{E}D, && (4.4)\\ \mathbb{E}D^2_{R,0} &= \mathbb{E}N(0,R)\,\sigma^2(D) + \mathbb{E}N(0,R)^2(\mathbb{E}D)^2, && (4.5)\end{aligned}$$

and since we assumed $D_{R,0}$ is gamma distributed

$$\mathbb{E}D^3_{R,0} = (1+c^2_{D_{R,0}})(1+2c^2_{D_{R,0}})(\mathbb{E}D_{R,0})^3. \tag{4.6}$$

Because zero is an arbitrary point in time and the moments of D are assumed to be given, the moments of $N(0,R)$ can be computed with formulas (2.20) and (2.21). These asymptotic relations are only valid when Condition $C.2$ ($R \geq Cond(A)$) holds. Note that relation (2.21) still contains the third moment of A. Following the same reasoning as before, we assume that A is gamma distributed. Then it is easy to see that the following relations hold

$$\begin{aligned} \mathbb{E}D_{R,0} &\approx \frac{R}{\mathbb{E}A}\mathbb{E}D, && (4.7)\\ \mathbb{E}D^2_{R,0} &\approx \Big(\frac{R^2}{(\mathbb{E}A)^2} + \frac{R}{\mathbb{E}A}(c^2_A + c^2_D) + \tfrac{1}{6}(1-c^4_A)\Big)\mathbb{E}D^2. && (4.8)\end{aligned}$$

From the fact that $\sigma(D_{R,0}) \geq 0$, we conclude from (4.8) that this approach is only valid when the coefficient of variation of the interarrival process satisfies Condition $C.3$ ($c^2_A \in [0, \frac{3R}{\mathbb{E}A} + \sqrt{\frac{9R^2}{(\mathbb{E}A)^2} + \frac{6R}{\mathbb{E}A}c^2_D + 1}]$). Hence, when $\mathbb{E}A$ is large with respect to R the region of application is restricted to situations for which $c^2_A \leq 1$. Substitution of the expressions (4.6), (4.7) and (4.8) into (4.2) and (4.3), completes the computation of the first two moments of $U_{R,0}$.

What remains to compute are the first two moments of $D(\tau_1, \tau_1 + L_1)$. Using relations (2.14) and (2.15) yields

$$\begin{aligned} \mathbb{E}D(\tau_1,\tau_1+L_1) &= \mathbb{E}N(\tau,\tau+L_1)\,\mathbb{E}D, && (4.9)\\ \mathbb{E}D(\tau_1,\tau_1+L_1)^2 &= \mathbb{E}N(\tau_1,\tau_1+L_1)\,\sigma^2(D) + \mathbb{E}N(\tau_1,\tau_1+L_1)^2(\mathbb{E}D)^2. && (4.10)\end{aligned}$$

Since τ_1 is an arbitrary moment in time, the moments of $N(\tau_1, \tau_1+L_1)$ can be approximated by its asymptotic relations (2.20) and (2.21) with t substituted by $\mathbb{E}L$, and t^2 substituted by $\mathbb{E}L^2$. The latter relations again only hold when Condition $C.4$ ($\mathbb{E}L \geq Cond(A)$) holds. Because we assumed that A is gamma distributed, it can be found that

$$\mathbb{E}D(\tau_1, \tau_1 + L_1) \approx \frac{\mathbb{E}L}{\mathbb{E}A}\mathbb{E}D, \tag{4.11}$$

$$\mathbb{E}D(\tau_1, \tau_1 + L_1)^2 \approx \Big(\frac{\mathbb{E}L^2}{(\mathbb{E}A)^2} + \frac{\mathbb{E}L}{\mathbb{E}A}(c_A^2 + c_D^2) + \tfrac{1}{6}(1 - c_A^4)\Big)(\mathbb{E}D)^2. \tag{4.12}$$

Analogous to Condition $C.3$ we conclude that (4.12) is only valid when Condition $C.5$ ($c_A^2 \in [0, \frac{3\mathbb{E}L}{\mathbb{E}A} + \sqrt{\frac{9(\mathbb{E}L)^2}{(\mathbb{E}A)^2} + \frac{6\sigma(L)^2}{(\mathbb{E}A)^2} + \frac{6\mathbb{E}L}{\mathbb{E}A}c_D^2 + 1}))$ is satisfied. In case $\mathbb{E}A$ is large with respect to $\mathbb{E}L$ the region of application is restricted to situations for which $c_A^2 \leq 1$. Thus using (4.2), (4.3), (4.11) and (4.12) we can find expressions for the first two moments of Z_1, which enables us to calculate the reorder point s.

We can distinguish between two kinds of conditions. The restrictions due to the use of asymptotic relations from renewal theory (Conditions $C.1$, $C.2$, and $C.4$), and the restrictions due the assumption that A is gamma distributed (Conditions $C.3$ and $C.5$). When Conditions $C.1$, $C.2$, or $C.4$ are violated, this means that we have to compute the renewal function $N(t)$ associated to the $\{A_i\}$ process for small values of t. In that case we suggest the method at the end of section 2.2. When Conditions $C.3$ or $C.5$ are violated (i.e., either $R \ll \mathbb{E}A$ or $\mathbb{E}L \ll \mathbb{E}A$) means that in case $R \ll \mathbb{E}A$, the frequency with which the inventory position is monitored is larger than for the continuous (s, Q) inventory model. Hence, a continuous inventory model (for example an (s, Q) or (s, S) policy) should be considered in these situations. For the situations in which $\mathbb{E}L \ll \mathbb{E}A$ there is a high probability of zero demand during the lead time. In that case we suggest, for situations that $c_A \approx 1$, the compound Bernoulli method described in Chapter 3, and otherwise we suggest to neglect lead times.

For the compound Poisson process, exact expressions for the moments of $N(0, R)$ and $N(\tau_1, \tau_1 + L_1)$ are available. Hence, we do not need the asymptotic relations (2.20) and (2.21), and therefore we do not require Conditions $C.2$ to $C.5$.

4.3 The discrete time model

In this section we describe the discrete time model, for which we assume that the time axis is divided into time units of equal length T, for example, days, and that R and the lead times L_k for $k = 1, 2, \ldots$ are integral multiple of T. Furthermore, we assume that the demand per time unit is registered for every time unit. Decisions about replenishments are made every R time units. The depletion of the inventory in the k-th time unit is equal to $D_{T,k}$, with $\mathbb{P}(D_{T,k} = 0) > 0$.

To obtain tractable results for the first two moments of Z_1 we have to assume that $\{D_{T,k}\}_{k=1}^{\infty}$ are independent and identically distributed random variables. We note here that, this assumption is often made in practice and in most text books without checking its validity.

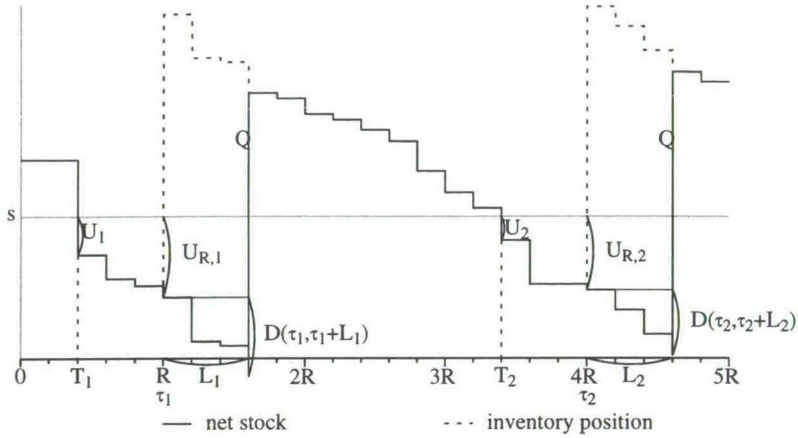


Figure 4.2: Evolution of the net stock and inventory position during the first replenishment cycle

Due to the transformation of the continuous time axis to a slotted time axis, events of several types may coincide in time with positive probability. Therefore, we have to specify the priority rule in which the events are handled. Note that different priority rules lead to different values for the reorder point. We assume that the depletion of stock $\{D_{T,k}\}_{k=1}^{\infty}$ is handled before a possible replenishment order at the end of a time unit. In Figure 4.2 the same sample path is used as in Figure 4.1, to illustrate the transformation to a slotted time axis and the impact of the priority rule. This priority rule, as can be seen in Figure 4.2, leads to a service which is lower than in case the service is measured continuously. Under this priority rule the demand during the lead time is given by

$$D(\tau_1, \tau_1 + L_1) = \sum_{j=\tau_1+1}^{\tau_1+L_1} D_{T,j} \tag{4.13}$$

Because the $\{D_{T,k}\}_{k=1}^{\infty}$ are non-negative i.i.d. random variables we can apply a well-known result for the first two moments of a stochastic number of i.i.d. random variables (see, for example, Hadley and Whitin(1963))

$$\begin{aligned} \mathbb{E}D(\tau_1, \tau_1 + L_1) &= \mathbb{E}L\, \mathbb{E}D_{T,1} && (4.14)\\ \mathbb{E}D(\tau_1, \tau_1 + L_1)^2 &= \mathbb{E}L\, \sigma^2(D_{T,1}) + \mathbb{E}L^2(\mathbb{E}D_{T,1})^2 && (4.15) \end{aligned}$$

Furthermore we have

$$D_{R,0} = \sum_{j=1}^{R} D_{T,j}, \tag{4.16}$$

yielding

$$\mathbb{E}D_{R,0} = R\,\mathbb{E}D_{T,1}, \tag{4.17}$$

$$\mathbb{E}D^2_{R,0} = R\,\sigma^2(D_{T,1}) + R^2(\mathbb{E}D_{T,1})^2. \tag{4.18}$$

To compute $\mathbb{E}D^3_{R,0}$ the same approach as in section 4.2 is followed, i.e., we assume that $D_{R,0}$ is gamma distributed. Then the third moment of $D_{R,0}$ is given by (4.6).

Since the $\{D_{T,k}\}_{k=1}^{\infty}$ are i.i.d. and the $\{D_{R,k}\}_{i=1}^{\infty}$ are disjunct collections of an identical number of $D_{T,k}$'s, it can be concluded that also the $\{D_{R,k}\}_{i=1}^{\infty}$ are i.i.d. The distribution function of $U_{R,1}$ can be approximated by the asymptotic forward recurrence time distribution of the renewal process generated by the sequence $\{D_{R,i}\}_{i=0}^{\infty}$. Using Theorem 2.1 for $k=1$ and $k=2$ yields

$$\mathbb{E}U_{R,1} \approx \frac{\mathbb{E}D^2_{R,0}}{2\mathbb{E}D_{R,0}}, \tag{4.19}$$

$$\mathbb{E}U^2_{R,1} \approx \frac{\mathbb{E}D^3_{R,0}}{3\mathbb{E}D_{R,0}}. \tag{4.20}$$

Note that relations (4.19) and (4.20) hold only when $Q \geq Cond(D_{R,0})$. Substitution of the two moments of $D_{T,1}$ and the first two moments of L in (4.14),(4.15) and (4.17) to (4.20) enables us to calculate the first two moments of Z_1. Next the distribution function of Z_1 is again approximated by an ME distribution (see section 2.4 and relation (4.1) is used to compute the reorder point s.

Note that Z_k for $k = 1, 2, \ldots$ can also be written as the undershoot U_k under s at T_k plus the demand during the pseudo lead time ($\hat{L}_k := L_k + \tau_k - T_k$). However, it can be proven that both approaches result in the same expressions for the first two moments of Z_k (see de Kok (1991c)).

For the special case of the compound Poisson process the $D_{T,k}$'s indeed are independent and identically distributed. In this situation the continuous time model and the discrete time model will lead to the same results.

4.4 Numerical results

In the simulation experiments we compare the reorder point s calculated for both models and the associated actual service level. We use a compound renewal process to model the demand process. More specifically the interarrival times and the demand size of a customer are i.i.d. random variables with ME distributions. The lead times are discrete random variables. In order to choose the appropriate discrete distribution function given the first two moments of the lead times, we use the method proposed by Adan et al. (1995) (see Chapter 3). Thus, in order to describe the inventory model we have to specify values for the first two moments of the interarrival times, demand sizes of customers and lead times, the length of the review period and of the time units, the replenishment quantity and the target service level ($P_{2,target}$). In Table 4.1 the parameters used in our experiments are given.

To obtain sufficient accurate values for the relevant input variables of the discrete time

Table 4.1: Input parameters for simulation experiments

Q	$P_{2,target}$	T	R	$\mathbb{E}A$	c_A	$(\mathbb{E}L, \sigma(L))$	$\mathbb{E}D$	$\sigma(D)$
50, 100	0.95	1	5	0.5; 1; 2	$\frac{1}{4}$; $\frac{1}{2}$; 1; $\frac{3}{2}$; 2; 3	(4,0); (10,2)	5	5

model (e.g. the first two moments of the demand per time unit), all the required input values are derived from a preceding simulation run. In operational settings this coincides with exact knowledge about the moments of the relevant random variables. The discrete time model requires the moments of $D_{T,1}$. The asymptotic expressions for the first two moments of $D_{T,1}$ are given by

$$\begin{aligned} \mathbb{E}D_{T,1} &\approx \mathbb{E}N(0,T)\,\mathbb{E}D \\ \mathbb{E}D_{T,1}^2 &\approx \mathbb{E}N(0,T)\,\sigma^2(D) + \mathbb{E}N(0,T)^2(\mathbb{E}D)^2. \end{aligned}$$

These expressions are only valid when $T \geq Cond(A)$. But, in our numerical experiments for several instances this restriction is not valid (e.g., $\mathbb{E}A = 1, c_A > 1$). We first simulated 1 subrun of 50.000 time units to obtain values for the moments of $D_{T,1}$ in order to make a fair comparison.

For each of the two methods a simulation is performed with the same seeds for the random generator. The reorder point s is calculated with the associated method. Then we simulated 10 times 50.000 time units. We denoted β_1 (β_2) as the actual P_2 service level computed by simulation for the associated value of the reorder point s, which was computed by the continuous time model (discrete time model).

First, we checked whether the conditions in 4.1 are valid. The results are tabulated in Tables 4.A.1 and 4.A.2 of Appendix 4.A. We see that the continuous time model performs excellent in every situation where no conditions are violated. When at least one of the conditions 4.1 is violated we note discrepancies between the target and actual service level. Especially when Conditions $C.2$ and $C.4$, concerning the computation of the renewal function $N(t)$ associated to the $\{A_i\}$ process for small values of t, are violated s is not computed correctly. In that case we suggest the method at the end of section 2.2. Secondly, we investigated to what extent a continuous time demand process can be approximated by a discrete time process. The performance of the discrete time model in the continuous time situation depends heavily on the coefficient of variation of the interarrival times. When the coefficient of variation of the interarrival times is smaller than one, the discrete time model tends to overestimate the reorder point s. Whereas, for situations in which the interarrival times are erratic ($c_A > 1$) the discrete time model tends to underestimate the reorder point (see Figure 4.3 to 4.4). For $c_A < 1$ the demands in two consecutive periods are negatively correlated. Consider for example the situation that $\mathbb{E}A = 2$ and $c_A = 0$, then negative correlation is obvious; when the demand in a certain period is positive then the demand in the next period is zero, and visa versa. For $c_A > 1$ it can be argued that the demands in

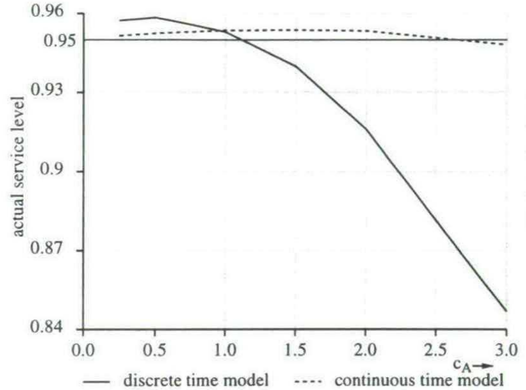


Figure 4.3: Actual service levels in case $(\mathbb{E}L, \sigma(L)) = (10, 2), \mathbb{E}A = 1$, $Q = 50$

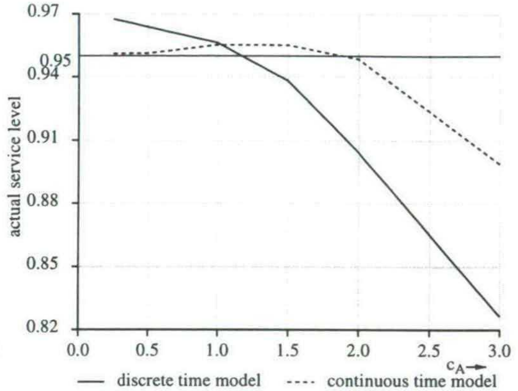


Figure 4.4: Actual service levels in case $(\mathbb{E}L, \sigma(L)) = (4, 0)$, $\mathbb{E}A = 2$, $Q = 50$

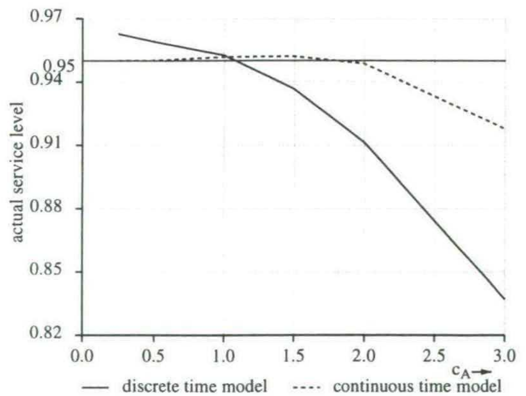


Figure 4.5: Actual service levels in case $(\mathbb{E}L, \sigma(L)) = (10, 2), \mathbb{E}A = 2$, $Q = 100$

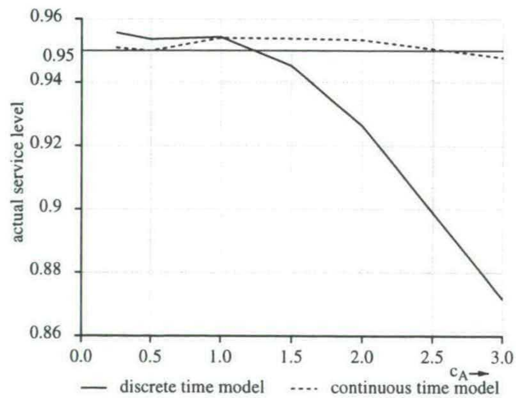


Figure 4.6: Actual service levels in case $(\mathbb{E}L, \sigma(L)) = (4, 0)$, $\mathbb{E}A = 1$, $Q = 100$

two consecutive periods are positively correlated. Only when $c_A = 1$, which represent the compound Poisson process, the D_n's are i.i.d. This means that the discrete time model is only valid when the demand process is a compound Poisson case. For situations where $c_A \neq 1$ the discrete time model is not valid. Even when the moments of the demand per time unit and the lead time are known exactly the method is not valid. Although, the deviations remain within reasonable bounds when $0.7 < c_A < 1.3$.

4.5 Conclusions and future research

In this chapter we compared two methods for the determination of the reorder point s in a (R, s, Q) inventory model subject to a P_2 service level constraint when the demand is described by a compound renewal process. The two methods differ in the modelling assumptions of the demand process, and therefore require different levels of information to feed the inventory models.

The compound renewal demand process is a versatile model to describe real-life demand processes. The demand process mostly used in literature is based on i.i.d. demand per period. In many practical cases, e.g. manufactures of components with a small number of customers, this is not a justified model. We have shown by discrete event simulation, that indeed this may result in very large errors in the P_2 service level. The continuous time model is robust within the area indicated by conditions 4.1. However, conditions $C.2$ and $C.4$ seem to be most severe for the quality of the fill rate performance.

The choice of which method to use in practical situations should be based also on the quality of the information available. The choice must be based on the coefficient of variation of the interarrival times. When detailed customer information is available the continuous time model should be used, at least within its applicability region. For future research we would like to point out two extensions. Firstly, when using such models in practice, estimates for the moments of the relevant stochastic variables are inevitable. Hence, the impact of the quality of the moment estimations on the performance of the methods should be investigated (see also Vaughan (1995)). Secondly, instead of using estimates for the moments of the variables one could also integrate forecasting procedures (e.g. exponential smoothing methods) directly in inventory models.

Appendix 4.A: Results of numerical experiments

Table 4.A.1 ($T = 1, R = 5, P_{2,target} = 0.95, (\mathbb{E}D\,;\sigma_D) = (5\,;5)), (\mathbb{E}L\,;\sigma_L) = (10\,;2)$)

		CR-method							DT-method	
Q	$(\mathbb{E}A, c_A)$	s	β_1	C.1	C.2	C.3	C.4	C.5	s	β_2
50	(0.5,0.25)	170.8	0.9540 (± 0.0005)	0	1	1	1	1	172.1	0.9564 (± 0.0020)
50	(0.5,0.50)	174.3	0.9538 (± 0.0004)	0	1	1	1	1	175.2	0.9560 (± 0.0020)
50	(0.5,1.00)	187.1	0.9536 (± 0.0006)	0	1	1	1	1	186.3	0.9523 (± 0.0026)
50	(0.5,1.50)	206.1	0.9532 (± 0.0010)	1	1	1	1	1	201.6	0.9460 (± 0.0044)
50	(0.5,2.00)	229.6	0.9539 (± 0.0009)	1	1	1	1	1	214.5	0.9328 (± 0.0052)
50	(0.5,3.00)	283.8	0.9541 (± 0.0012)	1	0	1	1	1	234.6	0.8914 (± 0.0066)
50	(1.0,0.25)	86.7	0.9515 (± 0.0010)	1	1	1	1	1	88.6	0.9573 (± 0.0034)
50	(1.0,0.50)	89.5	0.9524 (± 0.0008)	1	1	1	1	1	91.6	0.9584 (± 0.0026)
50	(1.0,1.00)	99.8	0.9536 (± 0.0012)	1	1	1	1	1	99.5	0.9530 (± 0.0038)
50	(1.0,1.50)	115.0	0.9537 (± 0.0012)	1	1	1	1	1	108.5	0.9398 (± 0.0048)
50	(1.0,2.00)	133.0	0.9534 (± 0.0013)	1	0	1	1	1	115.5	0.9160 (± 0.0060)
50	(1.0,3.00)	170.0	0.9481 (± 0.0009)	0	0	1	0	1	123.6	0.8469 (± 0.0052)
50	(2.0,0.25)	44.5	0.9521 (± 0.0007)	1	1	1	1	1	49.3	0.9697 (± 0.0022)
50	(2.0,0.50)	46.7	0.9518 (± 0.0009)	1	1	1	1	1	50.1	0.9642 (± 0.0032)
50	(2.0,1.00)	54.5	0.9545 (± 0.0010)	1	1	1	1	1	54.9	0.9564 (± 0.0030)
50	(2.0,1.50)	65.8	0.9532 (± 0.0013)	1	0	1	1	1	59.4	0.9330 (± 0.0052)
50	(2.0,2.00)	78.0	0.9503 (± 0.0010)	1	0	1	0	1	62.3	0.9007 (± 0.0058)
50	(2.0,3.00)	92.6	0.9170 (± 0.0012)	1	0	1	0	1	65.9	0.8085 (± 0.0068)
100	(0.5,0.25)	156.3	0.9521 (± 0.0006)	1	1	1	1	1	157.3	0.9542 (± 0.0022)
100	(0.5,0.50)	159.3	0.9518 (± 0.0006)	1	1	1	1	1	160.2	0.9538 (± 0.0020)
100	(0.5,1.00)	170.8	0.9526 (± 0.0009)	1	1	1	1	1	170.1	0.9515 (± 0.0032)
100	(0.5,1.50)	188.5	0.9522 (± 0.0009)	1	1	1	1	1	184.2	0.9462 (± 0.0034)
100	(0.5,2.00)	210.7	0.9528 (± 0.0008)	1	1	1	1	1	196.3	0.9336 (± 0.0026)
100	(0.5,3.00)	263.2	0.9541 (± 0.0012)	1	0	1	1	1	215.5	0.8949 (± 0.0066)
100	(1.0,0.25)	75.7	0.9496 (± 0.0007)	1	1	1	1	1	77.2	0.9541 (± 0.0024)
100	(1.0,0.50)	77.9	0.9507 (± 0.0006)	1	1	1	1	1	79.7	0.9554 (± 0.0022)
100	(1.0,1.00)	86.5	0.9512 (± 0.0008)	1	1	1	1	1	86.2	0.9506 (± 0.0032)
100	(1.0,1.50)	99.7	0.9516 (± 0.0011)	1	1	1	1	1	94.0	0.9389 (± 0.0046)
100	(1.0,2.00)	116.0	0.9514 (± 0.0012)	1	0	1	1	1	100.1	0.9210 (± 0.0038)
100	(1.0,3.00)	150.7	0.9472 (± 0.0010)	1	0	1	0	1	107.4	0.8585 (± 0.0052)
100	(2.0,0.25)	36.2	0.9502 (± 0.0007)	1	1	1	1	1	39.8	0.9628 (± 0.0020)
100	(2.0,0.50)	37.8	0.9502 (± 0.0006)	1	1	1	1	1	40.3	0.9591 (± 0.0018)
100	(2.0,1.00)	43.8	0.9518 (± 0.0009)	1	1	1	1	1	44.1	0.9526 (± 0.0034)
100	(2.0,1.50)	52.9	0.9521 (± 0.0010)	1	0	1	1	1	47.6	0.9366 (± 0.0034)
100	(2.0,2.00)	63.3	0.9486 (± 0.0013)	1	0	1	0	1	50.0	0.9112 (± 0.0046)
100	(2.0,3.00)	76.0	0.9177 (± 0.0011)	1	0	1	0	1	53.0	0.8370 (± 0.0072)

The validity of conditions 4.1 is given by a 0-1 column, where 0 is used when a condition is violated and 1 when the condition is valid.

Table 4.A.2 ($T = 1, P_{2,target} = 0.95, R = 5, (\mathbb{E}D\,;\,\sigma_D) = (5\,;\,5)), (\mathbb{E}L\,;\,\sigma_L) = (4\,;\,0)$)

		CR-method							DT-method	
Q	$(\mathbb{E}A, c_A)$	s	β_1	C.1	C.2	C.3	C.4	C.5	s	β_2
50	(0.5,0.25)	91.7	0.9530 (± 0.0018)	0	1	1	1	1	93.3	0.9577 (± 0.0017)
50	(0.5,0.50)	94.4	0.9528 (± 0.0026)	0	1	1	1	1	95.8	0.9573 (± 0.0018)
50	(0.5,1.00)	104.7	0.9566 (± 0.0018)	0	1	1	1	1	106.0	0.9591 (± 0.0020)
50	(0.5,1.50)	120.2	0.9580 (± 0.0029)	1	1	1	1	1	117.5	0.9532 (± 0.0033)
50	(0.5,2.00)	139.1	0.9559 (± 0.0039)	1	1	1	1	1	128.8	0.9373 (± 0.0024)
50	(0.5,3.00)	182.2	0.9560 (± 0.0015)	1	0	1	0	1	144.6	0.8936 (± 0.0020)
50	(1.0,0.25)	46.7	0.9513 (± 0.0028)	1	1	1	1	1	48.4	0.9575 (± 0.0023)
50	(1.0,0.50)	48.8	0.9518 (± 0.0020)	1	1	1	1	1	50.6	0.9573 (± 0.0015)
50	(1.0,1.00)	56.9	0.9563 (± 0.0029)	1	1	1	1	1	57.0	0.9568 (± 0.0032)
50	(1.0,1.50)	68.8	0.9564 (± 0.0038)	1	1	1	1	1	64.5	0.9432 (± 0.0037)
50	(1.0,2.00)	83.2	0.9561 (± 0.0047)	1	0	1	0	1	70.5	0.9224 (± 0.0032)
50	(1.0,3.00)	110.1	0.9502 (± 0.0016)	0	0	1	0	1	77.8	0.8562 (± 0.0026)
50	(2.0,0.25)	24.1	0.9511 (± 0.0032)	1	1	1	1	1	27.8	0.9676 (± 0.0028)
50	(2.0,0.50)	25.7	0.9513 (± 0.0031)	1	1	1	1	1	28.6	0.9638 (± 0.0027)
50	(2.0,1.00)	31.9	0.9553 (± 0.0048)	1	1	1	1	1	32.1	0.9563 (± 0.0046)
50	(2.0,1.50)	40.6	0.9552 (± 0.0051)	1	0	1	0	1	36.2	0.9382 (± 0.0059)
50	(2.0,2.00)	49.7	0.9485 (± 0.0028)	1	0	1	0	1	38.8	0.9042 (± 0.0074)
50	(2.0,3.00)	53.7	0.8988 (± 0.0028)	1	0	1	0	1	41.5	0.8264 (± 0.0037)
100	(0.5,0.25)	80.1	0.9515 (± 0.0022)	1	1	1	1	1	81.5	0.9559 (± 0.0024)
100	(0.5,0.50)	82.3	0.9530 (± 0.0019)	1	1	1	1	1	83.4	0.9563 (± 0.0022)
100	(0.5,1.00)	91.0	0.9532 (± 0.0018)	1	1	1	1	1	92.2	0.9557 (± 0.0016)
100	(0.5,1.50)	104.6	0.9554 (± 0.0040)	1	1	1	1	1	102.2	0.9514 (± 0.0042)
100	(0.5,2.00)	121.9	0.9551 (± 0.0019)	1	1	1	1	1	112.4	0.9398 (± 0.0033)
100	(0.5,3.00)	162.7	0.9554 (± 0.0018)	1	0	1	0	1	126.9	0.8974 (± 0.0020)
100	(1.0,0.25)	38.0	0.9510 (± 0.0019)	1	1	1	1	1	39.3	0.9557 (± 0.0025)
100	(1.0,0.50)	39.6	0.9500 (± 0.0007)	1	1	1	1	1	40.9	0.9536 (± 0.0012)
100	(1.0,1.00)	45.9	0.9540 (± 0.0036)	1	1	1	1	1	46.0	0.9543 (± 0.0037)
100	(1.0,1.50)	55.7	0.9538 (± 0.0026)	1	1	1	1	1	52.0	0.9452 (± 0.0039)
100	(1.0,2.00)	68.0	0.9534 (± 0.0042)	1	0	1	0	1	57.0	0.9262 (± 0.0042)
100	(1.0,3.00)	92.8	0.9478 (± 0.0011)	1	0	1	0	1	63.2	0.8717 (± 0.0023)
100	(2.0,0.25)	17.4	0.9504 (± 0.0024)	1	1	1	1	1	20.0	0.9619 (± 0.0026)
100	(2.0,0.50)	18.5	0.9518 (± 0.0025)	1	1	1	1	1	20.6	0.9597 (± 0.0032)
100	(2.0,1.00)	23.0	0.9547 (± 0.0038)	1	1	1	1	1	23.0	0.9549 (± 0.0038)
100	(2.0,1.50)	29.7	0.9486 (± 0.0067)	1	0	1	0	1	26.2	0.9367 (± 0.0068)
100	(2.0,2.00)	37.1	0.9454 (± 0.0041)	1	0	1	0	1	28.2	0.9080 (± 0.0064)
100	(2.0,3.00)	40.6	0.9039 (± 0.0047)	1	0	1	0	1	30.3	0.8547 (± 0.0049)

Chapter 5

A two-supplier model

This chapter is based on Janssen and de Kok (1997). When setting up an inventory policy, first of all it has to be decided whether to source all replenishments from one supplier, or to divide the orders among two or more sources. Both single sourcing and multiple sourcing have advantages and disadvantages (see, for example, Fearon (1993)). The selection of suppliers heavily depends on the purchase price (in addition to possible discounts) and on the terms of delivery (including the lead time characteristics). Adapting to discounts or other supply agreements often implies that the timing and sizes of future replenishment orders are not based on actual inventory position, but on externalties. Hence the choice of adapting to discounts is a trade-off between purchase price and ordering flexibility. Often a choice is made for either a flexible but expensive supplier or a rigid but cheap supplier. Yet, it may be profitable to use two suppliers as follows. A rigid supplier is used to obtain discounts or a low purchase price for the majority of the purchase volume, while a flexible supplier is used to react to short term changes in demand. For example, the largest share of the purchase volume is purchased at a manufacturer, and the remaining part from a distributor or wholesaler.

In this chapter we consider such a multiple sourcing purchasing strategy. General supply agreements are made with the main supplier to deliver a fixed quantity Q every review period. At review epochs the inventory position is evaluated. When the inventory position is below the order-up-to level S, an order is placed at the second supplier such that the inventory position is raised to the order-up-to level. Both suppliers have deterministic lead times. In case both lead times are deterministic (not necessarily of equal length) the replenishment cycles of both suppliers can be synchronized. In this chapter we define a replenishment cycle as the time interval between two successive arrivals of replenishment orders of supplier 1. When the lead times of at least one of the suppliers is stochastic, the number of deliveries of supplier 2 during a replenishment cycle becomes stochastic. This clearly makes the analysis of this model much more complicated. To concentrate on the essence of this chapter: to investigate the profitability of using two suppliers; we do not consider stochastic lead time.

Note that this multiple sourcing strategy is a combination of a push system (the main supplier delivers every review period a predetermined quantity) and a pull system (the replenishment orders placed at the second supplier are governed by an (R, S) replenishment policy). When using more than one source one must decide how to divide the purchase volume. In this chapter we develop an algorithm for the determination of the decision parameters S and Q, such that the long-run expected average costs per review period (the sum of the holding, purchasing, and ordering costs) are minimized subject to a service level constraint. Note that Q determines the partitioning of the purchase volume.

In the literature much attention is paid to multiple sourcing models (see, for example, Sculli and Wu (1981), Hong and Hayya (1992), and Lau and Zhao (1993)). The main idea of order splitting is to reduce lead time uncertainties by splitting the replenishment orders over more than one supplier at each replenishment epoch. Hence, the order splitting strategy differs from the two-supplier strategy as defined above, in the sense that in order splitting each supplier is used every time a replenishment is placed, whereas in the two-supplier strategy the second supplier is only used when necessary.

Furthermore, the model presented in this chapter strongly resembles the periodic review control for the stochastic product recovery problem with remanufacturing and procurements, see Inderfurth (1996). The remanufacturing problem considers an inventory model in which customers may return products, which then are remanufactured or disposed. However, besides the remanufactured stream of products, it is also possible to replenish products from an external supplier. For a periodic review control a framework is given in Inderfurth (1996) to analyze the structure of optimal decision rules. Hence identifying the remanufactured stream of goods by the replenishments which are pushed in the system by the main supplier, and replenishments of the external supplier by the replenishments of the second supplier, yields the resemblance. Inderfurth focuses on the optimal policy without service level restriction. In this chapter we consider a fixed decision strategy for both suppliers and search for the optimal decision parameters subject to a service level constraint.

More generally, we might say that both models fit into the general framework of inventory models with "negative demand". For the model of Inderfurth the negative demand is generated by the remanufactured items, while in our model negative demand (with respect to the reordering process at the second supplier) occurs when during a review period the actual demand is less than Q. Another situation where modelling through negative demand seems essential is the situation where replenishment occurs through a fixed internal production process and external (emergency) orders.

This chapter is organized as follows. In section 5.1 the two-supplier model is defined in more detail, and a method is presented to compute the optimal decision parameters. In section 5.2 the proposed method is verified by a number of simulation experiments. For a number of situations the optimal values for the decision parameters are computed by the algorithm, and the shape of the total relevant cost function is analysed. In section 5.3 conclusions are presented, and directions for future research are indicated.

5.1 Model description

We address an inventory replenishment strategy which is a combination of a pull and a push system. We consider a discrete time model (see Chapter 4), i.e. the time axis is divided into time units (e.g. days), and the demands per time unit are assumed to be i.i.d. random variables. Review periods are an integral number of time units. We assume that the lead times of both suppliers involved are deterministic and an integral number of time units. Furthermore, it is assumed that customer orders are handled at the end of a day just before the replenishment orders are handled.

The main supplier, denoted as supplier 1, will deliver each review period a fixed quantity of size Q. When Q is larger than or equal to the average demand during a review period the system is not stable, in the sense that the inventory will blow up. Therefore, we restrict Q to be smaller than the expected demand during a review period.

Each review period the inventory position is monitored, in order to make a replenishment decision for the second supplier. When the inventory position, say x, is below the order-up-to level, denoted by S, an order of size $S - x$ is placed at supplier 2. The lead times of replenishment orders from supplier 2 are equal to L. The actual lead time of the first supplier is not relevant for the reordering decision concerning the second supplier. The reason for this is that the ordering decisions for supplier 2 are based on the inventory position and hence only the moments at which this inventory position is changed are relevant. Therefore we can, without loss of generality, choose the length of the lead time of supplier 1 also equal to L, implying that the arrivals of replenishment orders from the two suppliers coincide in time. In summary, at each review epoch first the inventory position is raised with size Q (because a replenishment order at supplier 1 is booked, which will arrive L periods later), and secondly the inventory position is compared with the order-up-to level S in order to make a replenishment decision for supplier 2.

Customer orders which cannot be delivered directly from stock will be backordered. As performance criterion the P_2-service measure is used. In determining the long-run expected average costs per review period (denoted by $C(S, Q)$), we take into account ordering, purchasing, and holding cost. The ordering costs are proportional to the number of replenishment orders, but independent of the size of a replenishment order. The purchasing costs are proportional to the size of a replenishment order. Both the ordering costs and the purchasing costs may depend on the supplier. The holding costs are proportional to the size of the physical stock level. However, in spite of the difference in purchasing costs of the products, all the items in stock are accounted at the same unit value (for example the market value). Note that in case one would like to differentiate between holding costs for products of different suppliers, a specification of the customers delivery rule is required (for example first deliver products with the largest purchase price). This choice would be relevant depending on the internal accounting system of the company.

In order to derive an expression for $C(S,Q)$, the following additinal definitions are given:

m_i	the purchase costs per unit at supplier i, $(i \in \{1,2\})$;
a_i	the ordering costs per order at supplier i, $(i \in \{1,2\})$;
b_4	the holding costs per unit per review period;
$\Pi(S,Q)$	the probability that an order is placed at supplier 2 during an arbitrary review period;
X_n	the inventory position at the n-th review period immediately after a replenishment order at supplier 2 is placed, if any;
W_n	$= X_n - S$.

Note that b_4 is based on the product market value and not on the purchase costs. In the latter situation one has to distinguish between holding costs of products purchased from supplier 1 and supplier 2. The expected total relevant cost during a replenishment cycle can now be written as

$$C(S,Q) = \begin{cases} b_4 B_4(S,0) + m_2 \mathbb{E}D(0,R) + a_2 & Q = 0 \\ b_4 B_4(S,Q) + m_1 Q + m_2 \mathbb{E}(D(0,R) - Q) + a_1 + \Pi(S,Q) a_2 & 0 < Q < \mathbb{E}D(0,R) \\ \infty & Q \geq \mathbb{E}D(0,R) \end{cases} \quad (5.1)$$

The problem can now be formulated as

$$\begin{aligned} (\mathcal{P}_1) \quad & \text{minimize} \quad C(S,Q) \\ & \text{s.t.} \quad P_2(S,Q) = P_{2,target}; \\ & \qquad 0 \leq Q < \mathbb{E}D(0,R),\ S \geq 0. \end{aligned}$$

Clearly, in order to solve $(\mathcal{P}_1)$ we need expressions for $P_2(S,Q)$, $B_4(S,Q)$, and $\Pi(S,Q)$. An important difference with the standard (R,S) inventory model is that in the two supplier model it is possible that the inventory position at a review period is larger than the order-up-to level S. Therefore we cannot apply the standard theory of regenerative processes at two successive review periods. Using the same kind of reasoning as is done in Hadley and Whitin (1963, page 177), to compute the time averages of the performance measures $(P_2(S,Q), B_4(S,Q)$, and $\Pi(S,Q))$ in an (s,Q) inventory system, we may restrict ourselves to derive expressions for the performance measures in an arbitrary replenishment cycle.

In Figure 5.1 we assume that zero is an arbitrary review moment in time. Furthermore, we consider the first complete replenishment cycle after zero. In Figure 5.1 we have chosen L such that $2R \leq L < 3R$. Hence, just before a replenishment decision is made at epoch 0 there are two outstanding orders at supplier 1 (each of size Q), and there are at most two outstanding orders at supplier 2 (denoted by Q_0 and Q_1). At the review moment zero the inventory position is raised with a size of Q due to replenishments of supplier 1, and since the inventory position still is below S, a replenishment order of size Q_2 is placed at supplier 2. Let I_1 be the net stock at the beginning of the tagged replenishment cycle (time period L), and I_2 the net stock at the end of the tagged replenishment cycle (time period $R+L$). By tracing the sample path of the inventory position from 0 to $R+L$,

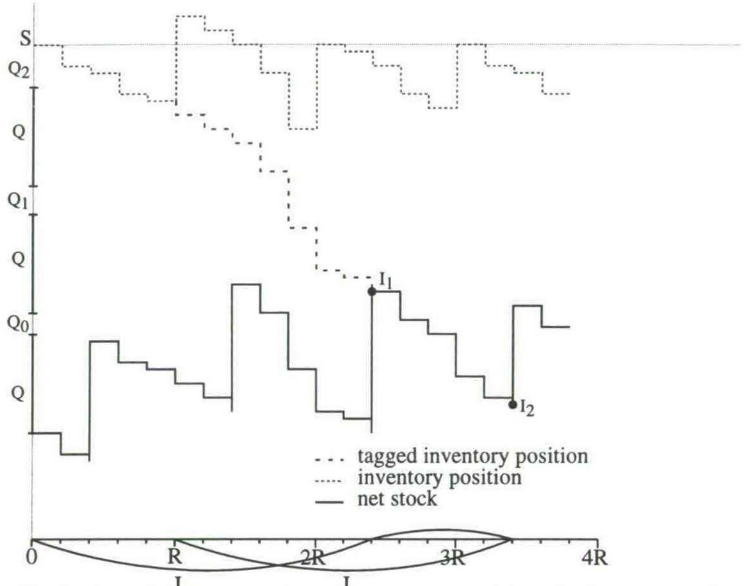


Figure 5.1: Evolution of the net stock and inventory position during a tagged replenishment cycle

neglecting all the replenishments that are made after time epoch 0, there is a clear relation between the inventory position just after time epoch 0 and the net stock at the beginning and end of the tagged replenishment cycle. Using the fact that the expected backlog at the beginning and at the end of the replenishment cycle are equal to $\mathbb{E}(-I_1)^+$ and $\mathbb{E}(-I_2)^+$ respectively, we can derive the following expression for the service level

$$P_2(S,Q) = 1 - \frac{\mathbb{E}(-I_2)^+ - \mathbb{E}(-I_1)^+}{\mathbb{E}(I_1 - I_2)}. \tag{5.2}$$

In order to derive expressions for $\mathbb{E}(-I_1)^+$ and $\mathbb{E}(-I_2)^+$ we have to distinguish between the situation $X_0 = S$ and $X_0 > S$. Then it is easy to see that

$$\begin{aligned}\mathbb{E}(-I_1)^+ &= \Pi(S,Q)\mathbb{E}((D(0,L) - X_0)^+ | X_0 = S) \\ &\quad + \big(1 - \Pi(S,Q)\big)\mathbb{E}((D(0,L) - X_0)^+ | X_0 > S)\end{aligned} \tag{5.3}$$

and

$$\begin{aligned}\mathbb{E}(-I_2)^+ &= \Pi(S,Q)\mathbb{E}((D(0,R+L) - X_0)^+ | X_0 = S) \\ &\quad + \big(1 - \Pi(S,Q)\big)\mathbb{E}((D(0,R+L) - X_0)^+ | X_0 > S).\end{aligned} \tag{5.4}$$

For $\mathbb{E}(I_1 - I_2)$, we find

$$\mathbb{E}(I_1 - I_2) = \mathbb{E}D(L, R+L). \tag{5.5}$$

By substituting (5.3-5.5) into (5.2), we find

$$1 - P_2(S,Q) = \Pi(S,Q)\frac{\mathbb{E}(D(0,R+L)-S)^+ - \mathbb{E}(D(0,L)-S)^+}{\mathbb{E}D(L,R+L)} \tag{5.6}$$
$$+ \big(1-\Pi(S,Q)\big)\frac{\mathbb{E}((D(0,R+L)-X_0)^+|X_0>S) - \mathbb{E}((D(0,L)-X_0)^+|X_0>S)}{\mathbb{E}D(L,R+L)}.$$

When the c.d.f. of $D(0,L), D(0,R+L)$ and $\mathbb{P}(X_0 \le x|X_0 > S)$ (the conditional distribution of X_0 given $X_0 > S$) are fitted on ME-distribution, see section 2.4, then (5.6) can be computed when (besides an expression for $\Pi(S,Q)$) expressions for $\mathbb{E}D(0,L)$, $\mathbb{E}D(0,L)^2$, $\mathbb{E}D(0,R+L)$, $\mathbb{E}D(0,R+L)^2$, $\mathbb{E}(X_0|X_0>S)$, and $\mathbb{E}(X_0^2|X_0>S)$ are available. From (2.14) and (2.15) we conclude

$$\mathbb{E}D(0,L) = L\mathbb{E}D, \tag{5.7}$$
$$\mathbb{E}D(0,L)^2 = L\sigma_D^2 + (L\,\mathbb{E}D)^2, \tag{5.8}$$

while analogous expressions exist for the first two moments of $D(0,R+L)$.

Consider the inventory positions at successive review epochs (immediately after a replenishment order at supplier 2 is placed, if any). Then it is easy to see that the following relation holds,

$$X_{n+1} = \max\{S, X_n + Q - D(nR,(n+1)R)\}. \tag{5.9}$$

Then using the relation between W_n and X_n gives

$$W_{n+1} = \max\{0, W_n + Q - D(nR,(n+1)R)\}. \tag{5.10}$$

Relation (5.10) is equivalent to the relation for the waiting times of two successive customers in a $GI|D|1$ queue with the distribution of the interarrival times equal to $F_{D(0,R)}(.)$ and with deterministic service time of length Q.

Chaudry (1992) gives an extensive overview of the available literature concerning the waiting times in a $GI|D|1$ queue. Most methods in literature require finding the roots of an equation (for example, in Chaudry (1992) the roots of equation $\tilde{A}(s)e^{-s/\mu} = 1$ are required, where $\tilde{A}(s)$ is the Laplace transform of the interarrival times of customers). Although this method is exact, we do not use this approach. The reason for this is that the method is relatively hard to implement. Instead, we use the approximate, however easy to implement, moment-iteration method for the waiting times in the $GI|G|1$ queue (see de Kok (1989)). This method computes values for $\mathbb{P}(W_0 > 0)$, $\mathbb{E}(W_0|W_0>0)$, and $\mathbb{E}(W_0^2|W_0>0)$. Note that $\mathbb{P}(W_0>0)$ is independent of S (see relation (5.10)). Using that $X_0 = W_0 + S$ we get the following relations

$$\Pi(S,Q) = 1 - \mathbb{P}(W_0>0); \tag{5.11}$$
$$\mathbb{E}(X_0|X_0>S) = \mathbb{E}(W_0|W_0>0) + S; \tag{5.12}$$
$$\mathbb{E}(X_0^2|X_0>S) = \mathbb{E}(W_0^2|W_0>0) + 2S\mathbb{E}(W_0|W_0>0) + S^2. \tag{5.13}$$

To obtain an expression for $B_4(S,Q)$ we again consider a tagged replenishment cycle. Note that lead times do not cross in time because they are deterministic. Therefore, all the outstanding orders, at the ordering epoch of the associated replenishment order of the tagged replenishment cycle, have arrived at the beginning of the tagged replenishment cycle. Hence the net stock (defined as the physical stock minus backlog) at the beginning of time epochs during the tagged replenishment cycle, $t \in \{L, R+L-1\}$, equals $X_0 - D(0,t)$. Then using again the renewal reward theorem, it is easy to see that the expected average physical stock is given by

$$\begin{aligned} B_4(S,Q) &= \frac{1}{R}\sum_{t=L}^{R+L-1} I\!E(X_0 - D(0,t))^+ \\ &= \Pi(S,Q)\frac{1}{R}\sum_{t=L}^{R+L-1} I\!E(S - D(0,t))^+ \\ &\quad + (1-\Pi(S,Q))\frac{1}{R}\sum_{t=L}^{R+L-1} I\!E(X_0 - D(0,t)|X_0 > S)^+. \end{aligned} \tag{5.14}$$

From (5.14) an explicit expression for $B_4(S,Q)$ can be obtained by fitting the conditional distribution of X_0 given $X_0 > S$ to an ME-distribution.

For an extensive exposition of the expected average physical stock in an (R,S) inventory model see, for example, de Kok (1991a). Now $C(S,Q)$ can be calculated for given values of S and Q. Note that for any given Q, the minimal value for S can be determined by solving $P_2(S,Q) = P_{2,target}$. Let $S^*(Q, P_{2,target})$ denote the optimal value of S as function of Q and $P_{2,target}$. Then $(\mathcal{P}_1)$ can be reformulated into the following one-dimensional optimization problem

$$\begin{aligned} (\mathcal{P}_2) \qquad \text{minimize} \quad & C(S^*(Q, P_{2,target}), Q) \\ & 0 \le Q < I\!ED(0,R). \end{aligned}$$

To solve $(\mathcal{P}_2)$ we use a local search procedure on the interval $(0, I\!ED(0,R))$ (see, for example Press et al. (1992)), and compared the solution with the single source situation: $C(S^*(0, P_{2,target}), 0)$. In our numerical investigations we did not find a counter example for the statement that the total relevant cost function is convex for $Q \in (0, I\!ED(0,R))$.

5.2 Numerical results

The values of the system and cost parameters for each of the three experiments are given in Table 5.1. The time unit is chosen to be a week, which means that the holding cost are based on a opportunity factor of 0.20 \$/\$/year. In the first experiment the quality is validated of the algorithm which is developed in the previous section. For given values of Q and $P_{2,target}$ the optimal value of S is computed by solving $P_2(S,Q) = P_{2,target}$. The actual service level for these values of Q, $P_{2,target}$ and S is computed by discrete event simulation and is denoted by $P_{2,sim}$. We simulated the system during 500.000 time units.

Table 5.1: Basic setting parameters for the experiments

	R	$\mathbb{E}D$	c_D	L	$P_{2,target}$	m_1	m_2	a_1	a_2	b_4
exp. 1	4	5	1	5	0.95	-	-	-	-	-
exp. 2	4	5	1	5	0.95	90	100	100	100	1.6
exp. 3	4	20	2	4	0.99	98	100	25	25	1.6

In Figure 5.2 we varied $P_{2,target}$ as $0.90, 0.91, \ldots, 0.99, 0.995, 0.999$, and fixed Q equal to 15. In Figure 5.3 we again varied $P_{2,target}$ as $0.90, 0.91, \ldots, 0.990, 0.995, 0.999$, and fixed Q equal to 15, however now we compared the expected average physical stock level computed by the algorithm (see formula (5.14)) with the average physical stock level computed by simulation. From both the experiments it is clear that the algorithm performs very well.

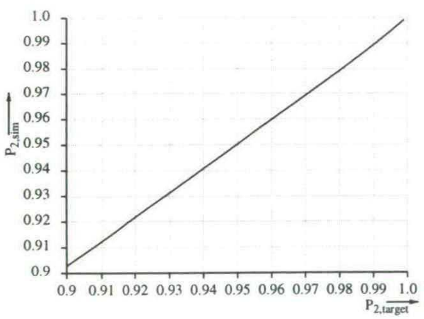


Figure 5.2: The comparison of $P_{2,target}$ with $P_{2,sim}$

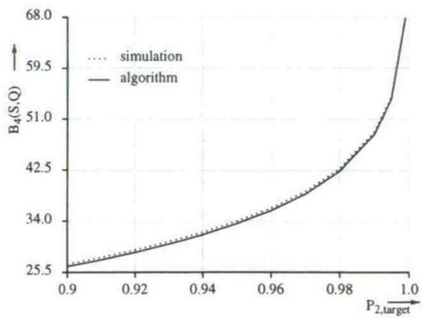


Figure 5.3: The comparison of the average physical stock levels computed by the algorithm and simulation respectively

In experiment 2 we check the conjecture that $C(S^*(Q, P_{2,target}), Q)$ is convex in Q. We computed $C(S^*(Q, P_{2,target}), Q)$, where Q is varied between 0 and $\mathbb{E}D(0, R)$. We define ρ as the fraction of the purchase volume delivered by supplier 1, that is $\rho := \frac{Q}{\mathbb{E}D(0,R)}$. In Figure 5.4 we considered $C(S^*(Q, P_{2,target}), Q)$ and varied b_4 between 0.4, 1.6 and 3.0 (\$/unit/review period). Note that $C(S^*(Q, P_{2,target}), Q)$ is very high for values of ρ almost equal to one. In this situation the load of the associated queuing system is near to one, which means long waiting lines. Therefore, the expected inventory position at the beginning of a cycle is very high, which implies high average stocks. This is also intuitively clear, since for values for Q almost equal to $\mathbb{E}D(0, R)$ there is no ordering flexibility, which means that in periods when demand is low unnecessary supplies are pushed into the system. Furthermore, we considered situations in which the holding costs are extremely low ($b_4 = 0.4$, representing a opportunity factor of 0.05 \$/\$/year) and extremely high ($b_4 = 3.0$, representing a opportunity factor of 0.40 \$/\$/year). Under this variation the optimal value for ρ does not vary much. This means that independent of the holding costs a large share of the replenishments should be purchased at supplier 1. Finally, this experiment indicates that the optimal value of ρ is rather high. The reason for this is the large difference between m_1 and m_2, which results in a purchase cost difference of \$ 200 per review period. Hence, the expected average physical stock has to differ at least with 125 units, which equals the demand of 5 review periods.

In the third experiment we investigated the structure of the optimal solution for the two-supplier problem. More precise, for given values of the input parameters we computed the optimal values for Q and S by solving $(\mathcal{P}_1)$. In Figure 5.5 we computed the optimal value of ρ where we varied m_1. Furthermore we varied the fixed costs a_1 and a_2 as follows: $\{a_1 = 200, a_2 = 100\}$; $\{a_1 = 0, a_2 = 0\}$; $\{a_1 = 100\,,\, a_2 = 200\}$ and $\{a_1 = 0\,,\, a_2 = 100\}$. Note that for $\{a_1 = 0, a_2 = 0\}$ the optimal value for ρ can take all values between 0 and 1, however only for small differences between m_1 and m_2 the optimal value for ρ is smaller than 0.8. When $a_1 > a_2$ the optimal value ρ drops to 0 before m_1 is equal to m_2, which means that only supplier 2 is used. Moreover, it is worthwhile to notice that when $a_1 < a_2$ it is even profitable to source products from supplier 1 even when $m_1 > m_2$. Hence, the increasing purchase cost is compensated by a decreasing ordering cost. In view

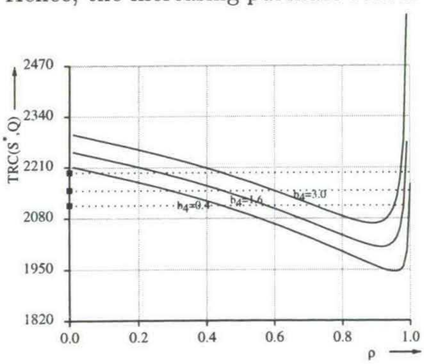


Figure 5.4: The total relevant costs as function of ρ

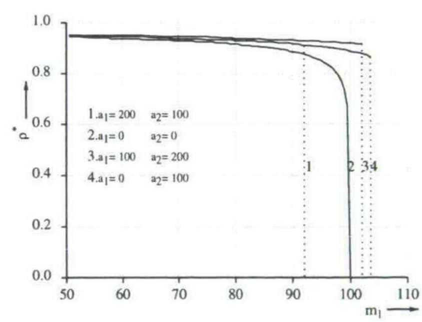


Figure 5.5: The optimal value of ρ as function of m_1

of the the structure of the optimal solution we then considered the situation where the variance of the customers demand sizes is high and the differences between m_1 and m_2 are small. In Figure 5.6 we again considered $C(S^*(Q, P_{2,target}), Q)$ and varied the coefficient of variation of the demand size (c_D) between 0.25, 1.00, and 2.00. The optimal value of ρ for $c_D = 0.25, 1.0$ and 2.0 is 0.980, 0.845 and 0.655 respectively. This indicates that the coefficient of variation of the demand size has a major impact on the value of the optimal ρ. For small values of c_D the optimal ρ is almost equal to one, and for large values of c_D the optimal value of ρ is relative small. This is to be expected, since one would use the flexible supplier whenever the demand process is erratic. In Figure 5.7 we computed the

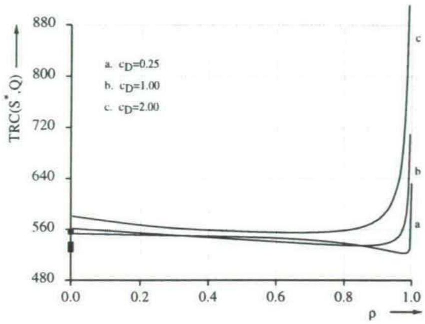


Figure 5.6: The total relevant costs as function of ρ

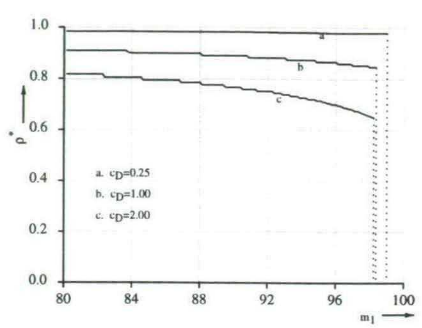


Figure 5.7: The optimal value of ρ as function of m_1

optimal value of ρ where we varied m_1 between 50 and 110. Furthermore we varied the coefficient of variation of the demand size (c_D) between 0.25, 1.00 and 2.00. Again, we can conclude that the c_D has a major impact on the optimal value of ρ.

5.3 Conclusions and future research

In this chapter we considered an inventory policy with two suppliers. A general supply agreement is made with one of the suppliers to deliver a fixed quantity Q every review period, whereas the replenishment decisions for the other supplier are governed by the (R, S) replenishment policy. Hence, when the inventory position at a review period is below the order-up-to level, S, an order is placed at supplier 2, such that the inventory position is raised to S. An algorithm is derived for the determination of the decision parameters S and Q for which the total relevant costs are minimized subject to a service level constraint.

Through comparisons with simulation results, the algorithm developed in this article appeared to perform excellently for all the experiments we considered. Furthermore, the numerical results showed the effectiveness and profitability of the multiple sourcing strategy as compared with the single sourcing strategy. It is clear that the profitability depends on the ratios between the ordering costs and purchase costs. However, the coefficient of variation of the demand turns out to be a crucial factor for the optimal value of Q. Only for large values of c_D it is profitable to purchase a large share of the purchase volume from the flexible and expensive supplier (supplier 2). This situation coincides with a Make-To-Order situation in which flexibility is very important. When the demand is stable (c_D is small) the need for flexibility is low and therefore the purchase costs dominate.

Although no numerical counter examples where found for the conjecture that the total relevant cost function is not convex in Q (neglecting the discontinuity at zero), a rigorous proof is needed to justify that $C(S^*(Q, P_{2,target}), Q)$ is indeed convex.

Several extensions are worthwhile to be considered. The generalization to stochastic lead times (with the non-overtaking restriction) will lead to complex expressions for the service equation. When the replenishment orders of supplier 2 are not restricted to arrive within a certain review period, the number of replenishments of supplier 1 within a replenishment cycle will not be fixed, which further complicates the determination of the service equation.

Another important extension is to allow a stochastic replenishment quantity (i.e. random yield) for supplier 1. The analysis is straightforward, when using the approximate methods, presented in this chapter, for obtaining values for the inventory position at review epochs.

Furthermore, other inventory control systems could be used for supplier 2. However, the key problem will be the determination of the inventory position at the possible replenishment epochs for the second supplier. For periodic review strategies this will again be possible using the approach presented in this chapter; for continuous review replenishment policies the way to go is less clear.

Chapter 6

Order splitting

During the last decade, order splitting has become a well-established issue in inventory literature. This vendor management strategy refers to the option of splitting a replenishment order among more than one supplier. The most important advantage of order splitting is the reduction in lead time uncertainties.

An important characteristic in analysing the performance of the order splitting concept is the time between the placement of an order and the arrival of the first partial delivery (the effective lead time). In Sculli and Wu (1981), Kelle and Silver (1990a, 1990b), and Guo and Ganeshan (1995) the theory of order statistics is used to derive analytical expressions for some characteristics of the first arriving partial delivery. In all these papers it is assumed that the demand rate is constant over time. The focus is on the optimal number of suppliers, which is determined based on the reduction in the safety stock.

Other papers focus on the decrease of the inventory holding cost due to the delayed replenishments (see, for example, Zhao and Lau (1992), Lau and Zhao (1993), Lau and Lau (1994), and Chiang and Chiang (1996)). In these papers the number of suppliers is mostly restricted to two, but demand is assumed to be stochastic. The papers focus either on minimizing the sum of holding, ordering, and shortage costs, or on minimizing the sum of ordering and holding costs subject to a service level constraint.

It has been shown that the profitability of order splitting depends on the ratio between the inventory holding cost and the extra transportation or ordering cost in case more than one supplier is used (see, for example, Larson (1989), Ramasesh et al. (1991), and Hong and Hayya (1992)).

In this chapter we consider an (s, Q) replenishment policy in which a replenishment order is split equally among n suppliers. In section 6.1 we will give a general model description for this order splitting model. Section 6.2 deals with the computational aspects of the performance measures in an order splitting model. In the next sections we will derive good approximations for the most common performance measures, merely by extending results from the standard (s, Q) inventory model. These expressions are used to determine optimal values for the control parameters: the reorder point, s, the reorder quantity, Q, and the number of suppliers, n. In section 6.3 we will discuss the use of the P_1 service measure in an order splitting model. In section 6.4 expressions for the P_2 and P_3 service measure are derived, which are validated in section 6.4.1. In section 6.5 we use a B_3 cost criterion to find optimal values of the control parameters s, Q and n.

6.1 Model description

In this single echelon inventory model with order splitting we assume that the demand process is a compound renewal process. That is, the interarrival times of customers can be described by the sequence $\{A_i\}_{i=1}^{\infty}$ of independent and identically distributed (i.i.d.) random variables with a common distribution function F_A, where A_i represents the time between the arrival of the i-th and $(i-1)$-th customer after time zero. Further, let $\tilde{A}$ be the asymptotic forward recurrence time associated with the sequence $\{A_i\}_{i=1}^{\infty}$. The demand sizes of the customers are described by the sequence $\{D_i\}_{i=1}^{\infty}$ of i.i.d. random variables with a common distribution function F_D, where D_i represents the demand size of the i-th customer after time zero. The sequence $\{D_i\}_{i=1}^{\infty}$ is independent of $\{A_i\}_{i=1}^{\infty}$.

Shortages are backordered, and replenishment decisions are based on the inventory position. The replenishment strategy that is considered is the continuous review (s, Q) policy. We assume that at most one time Q is ordered at a time. A replenishment order is equally split among n different suppliers. The suppliers have independent and identically distributed lead times with a common distribution function G. The subsequent realisations of the lead times of the n partial deliveries ranked in an increasing order are denoted by $L_{1:n} \leq L_{2:n} \leq \ldots \leq L_{n:n}$, with distribution functions $G_{k:n}$ for $k = 1, \ldots, n$ (the order statistics). It is assumed that deliveries of two successive replenishment orders (each consisting of n partial deliveries) do not cross in time. Thus, the last partial delivery of a replenishment order arrives before the first partial delivery of a subsequent replenishment order.

Renewal theory (see section 2.2) enables us to derive expressions for the long-run performance measures by deriving the related performance measures derived for an arbitrary replenishment cycle, being defined as the time between two subsequent epochs at which the last delivery of an order comes in.

Let zero be an arbitrary moment in time, and denote the j-th ordering epoch after zero by σ_j. Let $D(t_1, t_2)$ be the total demand during $(t_1, t_2]$, U_j the undershoot under s at σ_j. $L_{k:n}^{(j)}$ denotes the lead time of the k-th partial delivery in the j-th replenishment cycle after zero. Consider the first complete replenishment cycle after zero (see Figure 6.1). Denote for $k \in \{1, 2, \ldots, n\}$, I_k^b as the net stock at the beginning of the k-th sub-cycle in the first complete replenishment cycle after zero (just after the k-th partial delivery arrived), and I_k^e as the net stock at the end of the k-th sub-cycle in the first complete replenishment cycle (just before the k-th partial delivery arrives). Then it can be seen that (see Figure 6.1):

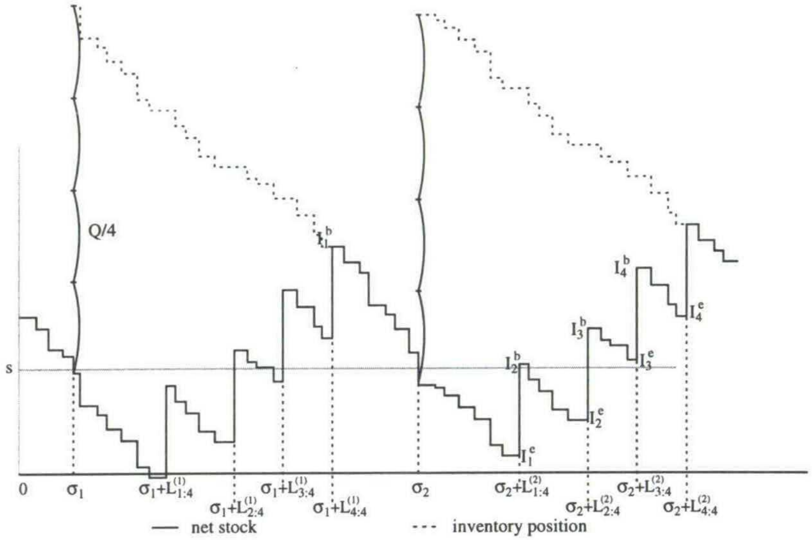


Figure 6.1: Evolution of the net stock and inventory position during a replenishment cycle for $n = 4$.

$$
\begin{aligned}
I_1^b &= s - U_1 + Q - D(\sigma_1, \sigma_1 + L^{(1)}_{n:n});\\
I_1^e &= s - U_2 - D(\sigma_2, \sigma_2 + L^{(2)}_{1:n});\\
I_k^b &= s - U_2 + \tfrac{k-1}{n} Q - D(\sigma_2, \sigma_2 + L^{(2)}_{k-1:n}), \quad k \in \{2, 3, \ldots, n\};\\
I_k^e &= s - U_2 + \tfrac{k-1}{n} Q - D(\sigma_2, \sigma_2 + L^{(2)}_{k:n}), \quad k \in \{2, 3, \ldots, n\}.
\end{aligned}
$$

Since the demand process is a compound renewal process and the lead times are i.i.d., it can be seen that $U_1 \overset{d}{=} U_2 \overset{d}{=} U$, and $D(\sigma_1, \sigma_1 + L^{(1)}_{n:n}) \overset{d}{=} D(\sigma_2, \sigma_2 + L^{(2)}_{n:n})$. Hence,

$$I_1^b \overset{d}{=} s - U_2 + Q - D(\sigma_2, \sigma_2 + L^{(2)}_{n:n}).$$

In the sequel we will suppress the index 2 in σ_2, U_2 and $L^{(2)}_{k:n}$.

6.2 Computational aspects

An important problem in inventory theory is to find the distribution of $D(\sigma, \sigma + L_{k:n})$ $(k = 1, \ldots, n)$ and the distribution of the undershoot U. In general these distribution functions are hard to obtain from F_A, F_D and G. To avoid this problem, we assume that $D(\sigma, \sigma + L_{k:n}) + U$ and $D(\sigma, \sigma + L_{k:n})$ $(k = 1, \ldots, n)$ are ME distributed. Since U is independent of $D(\sigma, \sigma + L_{k:n})$ it is sufficient to find expressions for the moments of U and $D(\sigma, \sigma + L_{k:n})$ separately.

Using results from section 2.2 gives

$$\mathbb{E}U \approx \frac{\mathbb{E}D^2}{2\mathbb{E}D}, \tag{6.1}$$

$$\mathbb{E}U^2 \approx \frac{\mathbb{E}D^3}{3\mathbb{E}D}, \tag{6.2}$$

and the first two moments of $D(\sigma, \sigma + L_{k:n})$ are given by (c.f. (2.14) and (2.15))

$$\mathbb{E}D(\sigma, \sigma + L_{k:n}) = \mathbb{E}N(\sigma, \sigma + L_{k:n})\mathbb{E}D, \tag{6.3}$$

$$\mathbb{E}D^2(\sigma, \sigma + L_{k:n}) = \mathbb{E}N(\sigma, \sigma + L_{k:n})\sigma^2(D) + \mathbb{E}N^2(\sigma, \sigma + L_{k:n})(\mathbb{E}D)^2, \tag{6.4}$$

where $N(\sigma, \sigma + L_{k:n})$ denotes the number of customer arrivals during $(\sigma, \sigma + L_{k:n})$. Recall, that σ is an order epoch. Hence, a customer arrived at epoch σ. Therefore we can use the following asymptotic expressions analogously to expressions (2.17) and (2.18)

$$\mathbb{E}N(\sigma, \sigma + L_{k:n}) \simeq \frac{\mathbb{E}L_{k:n}}{\mathbb{E}A} + \frac{\mathbb{E}A^2}{2\mathbb{E}A} - 1 \tag{6.5}$$

$$\mathbb{E}N^2(\sigma, \sigma + L_{k:n}) \simeq \frac{\mathbb{E}L_{k:n}^2}{(\mathbb{E}A)^2} + \mathbb{E}L_{k:n}\Big(\frac{2\mathbb{E}A^2}{(\mathbb{E}A)^3} - \frac{3}{\mathbb{E}A}\Big) + \frac{3(\mathbb{E}A^2)^2}{2(\mathbb{E}A)^4} - \frac{2\mathbb{E}A^3}{3(\mathbb{E}A)^3} - \frac{3\mathbb{E}A^2}{2(\mathbb{E}A)^2} + 1 \tag{6.6}$$

These asymptotic relations hold for $k = 1, \ldots, n$ only when $\mathbb{P}(L_{k:n} \leq A)$ is small, see section 2.2. What remains to compute are the moments of the order statistics $L_{k:n}$. Using a similar approach as is described in Balakrishnan and Cohen (1991, page 44), enables us to compute $\mathbb{E}L_{k:n}^m$ for $m \in \mathbb{N}$, and $k = 1, \ldots, n$, in case G is ME distributed (see Appendix 6.A).

6.3 On the P_1 service measure

In this section we discuss the non-stockout probability during a replenishment cycle (the P_1-service measure). The P_1 service measure is often considered in the order splitting literature (Sculli and Wu(1981), Kelle and Silver (1990a, 1990b), and Lau and Zhao (1993)). Since replenishment orders consist of n partial deliveries, we have to specify more precisely what we mean by the P_1 service measure.

Recall that a replenishment cycle is defined as the time between two last arrivals of partial deliveries of two successive replenishment orders. A natural way to extend the definition of the P_1 service measure to the multiple supplier model, is to consider only the first arriving partial delivery. In that case the P_1 measure is equal to the non-stockout probability just before the first arriving partial delivery. However, Kelle and Silver (1990b) calculated for a wide range of parameter values for Weibull-distributed lead times and constant demand the probability of a stockout occurrence just before the arrival of the last partial delivery. This probability turned out to be larger than the probability of a stockout occurrence just before any other partial delivery. In fact, for low values of Q the probability of a stockout occurrence before the last partial delivery may be the most critical. This makes the definition above for the P_1 service measure disputable. Alternatively, one could compute the non-stockout probabilities before each of the n partial deliveries, and then select that cycle with the lowest non-stockout probability as the critical one.

The most natural extension of the definition of the P_1 measure for multiple sourcing models is to define the P_1 service measure as one minus the probability that for at least one partial delivery a stockout occurs. In this section we will derive expressions for the stockout probabilities just before the k-th partial delivery ($k \in \{1, \ldots, n\}$), and the non-stockout probability during a complete replenishment cycle.

A disadvantage of the P_1 service measure in a single sourcing environment is that this measure is independent of Q, and therefore independent of the length of a replenishment cycle. Since high stockout probabilities during short intervals are much worse than high stockout probabilities during large intervals, and since a replenishment cycle in an order splitting environment is divided into n sub-cycles of different length, the probabilities of stockout during each of the sub-cycles must be weighted differently.

We are interested in the following performance measures:

$P_{1,all}(s, Q, n)$ the non-stockout probability during a complete replenishment cycle;

$P_{1,k}(s, Q, n)$ the non-stockout probability just before the k-th partial delivery ($k \in \{1, \ldots, n\}$);

$P_{1,min}(s, Q, n)$ the lowest non-stockout probability just before any of the n partial deliveries.

The non-stockout probability just before the k-th partial delivery ($k \in \{1, \ldots, n\}$) is then given by

$$\begin{aligned} P_{1,k}(s, Q, n) &= I\!P(I_k^e > 0) \\ &= I\!P(D(\sigma_2, \sigma_2 + L_{k:n}^{(2)}) + U_2 < s + \frac{k-1}{n} Q)). \end{aligned} \tag{6.7}$$

This formula differs from the formulas presented by Kelle and Silver (1990b) because they consider constant demand and therefore do not incorporate the undershoot.

The non-stockout probability just before the first and last arriving partial delivery is given by $P_{1,1}(s, Q, n)$ and $P_{1,n}(s, Q, n)$ respectively. Kelle and Silver (1990b) showed by numerical investigations that for Weibull-distributed lead times and constant demand $P_{1,n}(s, Q, n) \leq P_{1,k}(s, Q, n)$, $\forall k \in \{1, \ldots, n\}$. This means that $P_{1,n}(s, Q, n)$ is the most critical non-stockout probability. But from our numerical examples it turned out that $P_{1,n}(s, Q, n) \leq P_{1,k}(s, Q, n)$ is not in general true $\forall k \in \{1, \ldots, n\}$. Therefore, it might be an appropriate extension to focus on the lowest non-stockout probability, i.e.

$$P_{1,min}(s, Q, n) = \min_k \{P_{1,k}(s, Q, n)\}. \tag{6.8}$$

This intuitively appealing service measure has not been investigated in literature before.

As has been argued in the introduction the most natural extension of the P_1 service measure for a multiple sourcing environment is the non-stockout probability during a complete replenishment cycle, i.e.

$$P_{1,all}(s,Q,n) \quad = \quad I\!P(I_1 > 0, \ldots, I_n > 0). \tag{6.9}$$

However, due to the correlation of stockouts just before two different partial deliveries, it will be very cumbersome computing values for this definition of P_1. In Lau and Zhao this correlation is neglected. They used for the stockout probability the expression

$$\sum_{k=1}^{n}(1 - I\!P(I_k > 0)), \tag{6.10}$$

which is actually the expected number of sub-cycles (as Lau and Zhao indicate in their paper) in which there is a stockout just before the partial delivery arrives. Clearly, this quantity lies within $[0, n]$. Namely, consider the situation that $s < -Q$, then $I_k < 0$ for $k = 1, \ldots, n$, so the expected number of sub-cycles in which there is a stockout equals n.

Also the expected number of sub-cycles in which there is a stockout just before the partial delivery arrives could be used as performance measures. The major disadvantage of this measure is, as for the P_1 measure, that the sub-cycles are of different length. This means that stockouts in the different sub-cycles should be weighted differently.

We can compute the probabilities $P_{1,k}$ for $k \in \{1, \ldots, n\}$ easily by using that $D(\sigma_2, \sigma_2 + L_{k:n}^{(2)}) + U_2$ is ME distributed, see section 6.2. Then the P_1 service level constraint can be used to compute the optimal value of s.

6.4 The P_2 and P_3 service measure

In this section we consider both the P_2 and P_3 service level constraint and is based on Janssen and de Kok (1997b). The P_2 service measure is well-studied and widely applied in practice. However, in the order splitting literature this service measure is almost unexplored. The P_2 service measure is only discussed in Chiang and Benton (1994) and Chiang and Chiang (1996). The expressions derived by Chiang and Chiang (1996) are based on a model with deterministic lead times and normally distributed demand. Chiang and Benton (1994) consider the case with two suppliers with shifted exponential lead times, and normally distributed demand.

The P_3 service measure finds common application in the case of equipment used for emergency purposes (Silver and Peterson (1985) page 265). In this section we will derive an expression for this measure in an order splitting environment. Moreover, this service measure naturally arises as optimality condition when minimizing the sum of the expected ordering, holding and shortage costs, where the shortage costs are proportional to the expected average backlog level (see section 2.3).

We are interested in the following long-run performance measures, cf. section 2.1.4.

$P_2(s, Q, n)$ the fill rate (the fraction of demand directly delivered from stock);

$P_3(s, Q, n)$ the fraction of the time the physical stock is positive.

First we will derive an expression for $P_2(s, Q, n)$. The expected shortage during the k-th sub-cycle is given by the expected shortage at the end of the sub-cycle minus the shortage at the beginning (the last term for avoiding double counting). And, since the demand and lead time process are stationary it follows that the expected demand during a complete replenishment cycle is equal to Q. Hence,

$$P_2(s, Q, n) = 1 - \frac{1}{Q} \sum_{k=1}^{n} \mathbb{E}(-I_k^e)^+ - \mathbb{E}(-I_k^b)^+. \tag{6.11}$$

For the $P_3(s, Q, n)$ service measure we have the following result (see also Theorem 2.3).

Theorem 6.1

$$\begin{aligned} P_3(s, Q, n) \approx\ & \frac{\mathbb{E}\tilde{A} - \mathbb{E}A}{\mathbb{E}A} \mathbb{E}D \sum_{k=1}^{n} \frac{\mathbb{P}(I_k^b < 0) - \mathbb{P}(I_k^e < 0)}{Q} \\ & + \sum_{k=1}^{n} \frac{\mathbb{E}(I_k^b + U)^+ - \mathbb{E}(I_k^e + U)^+}{Q}. \end{aligned} \tag{6.12}$$

Proof

Define $T(x)$ as the expected total time that the physical stock is positive, in case the physical stock level at epoch 0 equals x ($x \geq 0$), there are no outstanding replenishment orders, and time epoch 0 is an arrival epoch of a customer. Let M denote the renewal function associated with F_D, then

$$T(x) = \mathbb{E}A\ M(x) \tag{6.13}$$

Analogously, we define $\tilde{T}(x)$ as the expected total time that the physical stock is positive, in case the physical stock level on epoch 0 equals x ($x \geq 0$), there are no outstanding replenishment orders, and time epoch 0 is an arbitrary moment in time. Let $\tilde{A}$ be the arrival time of the first customer after zero. Conditioning on the first arriving customer after time epoch 0, results into

$$\tilde{T}(x) = (\mathbb{E}\tilde{A} - \mathbb{E}A) + \mathbb{E}A\, M(x) \tag{6.14}$$

Now, consider the k-th sub-cycle ($k \in \{1, \ldots, n\}$). The expected physical stock at the beginning of the k-th sub-cycle (just after the replenishment arrived) equals $(I_k^b)^+$, whereas the expected physical stock at the end of the k-th sub-cycle (just before the replenishment arrives) equals to $(I_k^e)^+$. Then it is easy to see that the expected time that the physical stock is positive during the k-th sub-cycle is given by $I\!E\tilde{T}((I_k^b)^+) - I\!E\tilde{T}((I_k^e)^+)$. By using relation (6.14), Lemma 2.2, and by conditioning on I_k^b, we find

$$\begin{aligned}
I\!E\tilde{T}((I_k^b)^+) &= \int_0^{s+\frac{k-1}{n}Q} \tilde{T}(s+\frac{k-1}{n}Q-x)dF_{D(\sigma,\sigma+L_{k-1:n})+U}(x) \\
&= (I\!E\tilde{A}-I\!EA)\int_0^{s+\frac{k-1}{n}Q} dF_{D(\sigma,\sigma+L_{k-1:n})+U}(x) \\
&\quad + I\!EA\int_0^{s+\frac{k-1}{n}Q}\int_0^{s+\frac{k-1}{n}Q-x} d(M*U)(y)dF_{D(\sigma,\sigma+L_{k-1:n})}(x) \\
&\approx (I\!E\tilde{A}-I\!EA)I\!P(I_k^b<0)+\frac{I\!EAI\!E((I_k^b+U)^+)}{I\!ED},
\end{aligned}$$

and for $I\!E\tilde{T}((I_k^b)^+)$ a similar expression can be derived. Finally, using that the length of a replenishment cycle equals $\frac{QEA}{ED}$, and summing up the expected time the net stock is positive during the n sub-cycles, yields

$$\begin{aligned}
P_3(s,Q,n) &= \frac{I\!ED}{QI\!EA}\sum_{k=1}^{n}\left(I\!E\tilde{T}((I_k^b)^+)-I\!E\tilde{T}((I_k^e)^+)\right) \\
&\approx \frac{I\!E\tilde{A}-I\!EA}{I\!EA}I\!ED\sum_{k=1}^{n}\frac{I\!P(I_k^b<0)-I\!P(I_k^e<0)}{Q} \\
&\quad +\sum_{k=1}^{n}\frac{I\!E(I_k^b+U)^+-I\!E(I_k^e+U)^+}{Q}
\end{aligned} \tag{6.15}$$

which completes the proof. □

Note that $I_k^e+U = s+\frac{k-1}{n}Q - D(\sigma,\sigma+L_{k:n})$. For situations in which the undershoot is negligible and the demand process is a compound Poisson process, it can be shown that $P_2(s,Q,n)$ and $P_3(s,Q,n)$ are equal.

6.4.1 Numerical validation

As has been argued in the introduction, the service level constraint can be used to determine the reorder point s for given values of Q and n. The appropriate value of s can be found by solving $P_i(s,Q,n) = P_{i,target}$. Since for $i=1,2$ $0 \leq P_i(s,Q,n) \leq 1$ and $P_i(s,Q,n)$ is increasing in s, the roots of the $P_i(s,Q,n) = P_{i,target}$ can be found simply by using a local search algorithm (see, for example, Press et al.(1992)).

In this section we validate the quality of the algorithm described in the previously section for computing $P_2(s,Q,n)$ and $P_3(s,Q,n)$ by discrete event simulation. The numerical experiments are performed for a wide range of parameter values. The input values of the system parameters are given in Table 6.1. Each of the 3240 experiments consist of 10 subruns of 100.000 time units (exclusive 1 initialisation run). The demands, customer arrivals

Table 6.1: Basic parameter setting for the numerical validation

n	$\mathbb{E}D$	c_D	$\mathbb{E}A$	c_A	$\mathbb{E}L$	c_L	Q	$P_{3,target}$
1	5	$\frac{1}{2}$	$\mathbb{E}D/5$	$\frac{1}{2}$	5	$\frac{3}{10}$	50	0.50
2	10	1		1	10	$\frac{1}{2}$	100	0.99
3		2		2		1	250	
5								
10								

and lead times are drawn from ME distributions. For given values of n, Q and $P_{3,target}$ the reorder point s^* was determined by solving $P_3(s, Q, n) = P_{3,target}$. The output of the simulation experiment are values for the service measures under consideration, denoted by $P_{2,sim}$ and $P_{3,sim}$ respectively, and the fraction of the partial deliveries that crossed any partial deliveries of previous placed replenishment orders, denoted by C_{sim}.

In Table 6.2 we summarized the results of these experiments. Each line in Table 6.2 represents 180 simulation experiments, and we calculated the mean absolute deviation of $P_{3,target}$ and $P_{3,sim}$ and the mean absolute deviation of $P_2(s^*, Q, n)$ (denoted by $P_{2,target}$) and $P_{2,sim}$.

From these experiments we can draw the following conclusions about the quality of the expressions for the performance measures computed by the proposed algorithm in this section:

- For the situations in which $Q = 100$ or $Q = 250$, the proposed algorithm performs very good in all cases that are considered. Both $|P_{3,target} - P_{3,sim}|$ and $|P_2 - P_{2,sim}|$ are small, even for high coefficients of variation.

- For the situations where $Q = 50$ and $c_L = 1$, we see discrepancies between the target and achieved P_3 service level. The explanation for this deviation is that a large fraction of partial deliveries does cross (up to 58%). When this occurs we need to reconsider the determination of the moments $\mathbb{E}(L^{(2)}_{k:n})^m$, which are now determined by realisations of lead times of partial deliveries from more than one replenishment cycles.

These results point out that the proposed algorithm performs very wel, except for cases where crossing of orders indeed occurs with high probability.

In addition to the previous numerical validation we compared our results with results from Chiang and Benton (1994) to check the performance of our algorithm under different model assumptions. Chiang and Benton (1994) considered an (s, Q) inventory model with two suppliers, shifted exponentially distributed lead times, and normally distributed demand. Chiang and Chiang (1996) and Chiang and Benton (1994) are the only two papers in literature that consider the P_2 service measure. But in both papers the undershoot of

Table 6.2: The absolute deviations of the values of P_2 and P_3 computed by the algorithm and simulation.

Q	$P_{3,target}$	c_L	$\lvert P_{3,target} - P_{3,sim}\rvert$	$\lvert P_{2,target} - P_{2,sim}\rvert$	C_{sim}
50	0.50	0.3	0.0143	0.0085	0.4889
50	0.99	0.3	0.0037	0.0047	0.4892
100	0.50	0.3	0.0032	0.0024	0.1446
100	0.99	0.3	0.0022	0.0022	0.1447
250	0.50	0.3	0.0013	0.0011	0.0048
250	0.99	0.3	0.0011	0.0008	0.0047
50	0.50	0.5	0.0331	0.0257	0.5376
50	0.99	0.5	0.0064	0.0084	0.5374
100	0.50	0.5	0.0072	0.0055	0.1987
100	0.99	0.5	0.0039	0.0043	0.1987
250	0.50	0.5	0.0015	0.0016	0.0085
250	0.99	0.5	0.0015	0.0010	0.0085
50	0.50	1.0	0.1307	0.0949	0.5878
50	0.99	1.0	0.0096	0.0118	0.5878
100	0.50	1.0	0.0422	0.0319	0.3094
100	0.99	1.0	0.0085	0.0098	0.3091
250	0.50	1.0	0.0032	0.0031	0.0377
250	0.99	1.0	0.0053	0.0027	0.0377

the reorder level at ordering epochs is neglected and double-counting is allowed of shortages just before two subsequent partial deliveries. It is easy to see that double-counting can lead to negative service levels. In the computational experiments of Chiang and Benton (1994) they consider normally distributed demand per day with mean 50 units/day and variance 10 units/day, and shifted-exponential lead times with mean 8 and variance equal to 4. To make a fair comparison we used the same first two moments for our models, i.e. $\mathbb{E}A = 1, c_A = 0, \mathbb{E}D = 50$, $c_D = 0.2$, $\mathbb{E}L = 8$, and $c_L = 0.5$. We simulated the model under the conditions of Chiang and Benton, that is with shifted-exponential lead times and normally distributed demand, and compared the results in Table 6.3.

The examples considered in Chiang and Benton (1994) are for very high service levels and for rather stable demand processes. Then neglecting of the undershoot and double-counting have no impact on the calculated reorder level, which is reflected by the good results in Table 6.3. In spite of the difference in the model assumption and the simulated distribution functions our method did perform very well. In case we simulated lead times from a ME distribution the model presented here performs slightly better than the results of Chiang and Benton. Hence the model presented in this chapter is applicable for many situations. This in contrast to the model of Chiang and Benton which is restrict to normally distributed demand and discrete time models.

Table 6.3: A comparison of results from Chiang and Benton (CB) with our model (JK)

Q	$P_{2,target}$		s	$P_{2,sim}$
1350	0.9952	(CB)	531	0.9953 ($\pm$ 0.0009)
		(JK)	521	0.9950 ($\pm$ 0.0009)
1600	0.9954	(CB)	493	0.9956 ($\pm$ 0.0005)
		(JK)	504	0.9960 ($\pm$ 0.0005)
2050	0.9956	(CB)	455	0.9956 ($\pm$ 0.0008)
		(JK)	482	0.9970 ($\pm$ 0.0007)
2850	0.9959	(CB)	422	0.9959 ($\pm$ 0.0008)
		(JK)	456	0.9971 ($\pm$ 0.0007)

6.5 The B_3 cost criterion

In this section the values of the control parameters s, Q and n are determined such that the total sum of long-run ordering, holding and backordering cost per unit time are minimized. This section is based on Janssen and de Kok (1997a).

Let $C(s,Q,n)$ denote the total of ordering, holding and backordering cost per unit time incurred during an arbitrary replenishment cycle. The holding costs are proportional to the expected average physical stock level ($B_4(s,Q,n)$), i.e. stocking one unit of product costs b_4 \$ per day. Hong and Hayya (1990) investigated the effects of the ordering costs on models with order splitting. In particular they considered ordering costs that depend on the number of suppliers (denoted by $A(n)$). They showed that the optimal number of suppliers is very sensitive to the shape of $A(n)$. We use the following simple function for the ordering costs,

$$A(n) = n^c K, \quad n \in \mathbb{N}, c \in \mathbb{R}, \tag{6.16}$$

where K is a fixed cost component, and c determines the shape of $A(n)$. By varying c we can model a convex, concave, or a linear ordering cost function. Backordering costs are proportional to the number of units short, which coincides with the B_3 criterion: each unit short is charged with an amount of say b_3 \$ per time unit. Hence,

$$C(s,Q,n) = \frac{A(n)}{T(s,Q,n)} + b_3 B_3(s,Q,n) + b_4 B_4(s,Q,n), \tag{6.17}$$

where (cf. section 2.1.4)

$T(s,Q,n)$ the expected length of an arbitrary replenishment cycle;
$B_3(s,Q,n)$ the average backlog level during an arbitrary replenishment cycle;
$B_4(s,Q,n)$ the average physical stock level during an arbitrary replenishment cycle.

As in the standard (s,Q) inventory model, it follows that the expected demand during a replenishment cycle is equal to Q. Then it is easy to see that

$$T(s,Q,n) = \frac{Q\mathbb{E}A}{\mathbb{E}D}. \tag{6.18}$$

Note that $T(s, Q, n)$ is independent of s and n.

In order to derive an expression for $B_4(s, Q, n)$ we need the expected surface between the physical stock level and the zero level during a replenishment cycle. Theorem 6.2 below provides an explicit expression for the $B_4(s, Q, n)$.

Theorem 6.2

$$B_4(s,Q,n) \approx \frac{\mathbb{E}\tilde{A} - \mathbb{E}A}{\mathbb{E}A}\mathbb{E}D\sum_{k=1}^{n}\frac{\mathbb{E}(I_k^b)^+ - \mathbb{E}(I_k^e)^+}{Q} + \sum_{k=1}^{n}\frac{\mathbb{E}((I_k^b+U)^+)^2 - \mathbb{E}((I_k^e+U)^+)^2}{2Q}. \tag{6.19}$$

Proof

Define $H(x)$ (and $\tilde{H}(x)$) as the expected area between the physical inventory level and the zero level, in case the physical stock level on epoch 0 equals x ($x \geq 0$), there are no outstanding replenishment orders, and time epoch 0 is an arrival moment of a customer (time epoch zero is an arbitrary moment in time, respectively). By conditioning on the first arriving customer after time epoch 0, we find

$$H(x) = x\mathbb{E}A + \int_0^x H(x-y)dF_D(y). \tag{6.20}$$

Let M be the renewal function associated with F_D, then writing out recurrence relation (6.20) yields

$$H(x) = \mathbb{E}A\int_0^x (x-y)dM(y) \tag{6.21}$$

For $\tilde{H}(x)$, conditioning on the first arriving customer after time epoch 0 results into

$$\tilde{H}(x) = x\mathbb{E}\tilde{A} + \int_0^x H(x-y)dF_D(y). \tag{6.22}$$

Using relations (6.20) and (6.21) gives

$$\tilde{H}(x) = (\mathbb{E}\tilde{A} - \mathbb{E}A)x + \mathbb{E}A\int_0^x (x-y)dM(y) \tag{6.23}$$

Now, consider the k-th sub-cycle ($k \in \{1,\ldots,n\}$). The physical stock at the beginning of the k-th sub-cycle (just after the replenishment arrived) is equal $(I_k^b)^+$, whereas the physical stock at the end of the k-th sub-cycle (just before the replenishment arrives) is equal to $(I_k^e)^+$. Then it is easy to see that the expected area between the physical inventory level and the zero level within the k-th sub-cycle is given by $I\!E\tilde{H}((I_k^b)^+) - I\!E\tilde{H}((I_k^e)^+)$. By using (6.23), Lemma 2.2, and by conditioning on I_k^b, we find

$$\begin{aligned}
I\!E\tilde{H}((I_k^b)^+) &= \int_0^{s+\frac{k-1}{n}Q} \tilde{H}(s+\frac{k-1}{n}Q-x)dF_{D(\sigma,\sigma+L_{k-1:n})+U}(x) \\
&= (I\!E\tilde{A}-I\!EA)\int_0^{s+\frac{k-1}{n}Q}(s+\frac{k-1}{n}Q-x)dF_{D(\sigma,\sigma+L_{k-1:n})+U}(x) \\
&\quad + \frac{I\!EA}{2I\!ED}\int_0^{s+\frac{k-1}{n}Q}(s+\frac{k-1}{n}Q-x)^2 dF_{D(\sigma,\sigma+L_{k-1:n})}(x) \\
&\approx (I\!E\tilde{A}-I\!EA)I\!E((I_k^b)^+) + \frac{I\!EAI\!E((I_k^b+U)^+)^2}{2I\!ED}
\end{aligned}$$

and for $I\!E\tilde{H}((I_k^b)^+)$ an analogue expression holds.

Finally using that the length of a replenishment cycle equals $\frac{QEA}{ED}$ and summing up the expected area's of the n sub-cycles, yields

$$\begin{aligned}
B_4(s,Q,n) &= \sum_{k=1}^{n}\left(I\!E\tilde{H}((I_k^b)^+) - I\!E\tilde{H}((I_k^e)^+)\right) \\
&\approx \frac{I\!E\tilde{A}-I\!EA}{I\!EA}I\!ED\sum_{k=1}^{n}\frac{I\!E(I_k^b)^+ - I\!E(I_k^e)^+}{Q} \\
&\quad + \sum_{k=1}^{n}\frac{I\!E((I_k^b+U)^+)^2 - I\!E((I_k^e+U)^+)^2}{2Q},
\end{aligned} \tag{6.24}$$

which completes the proof of relation 6.19. □

Theorem 6.3

$$\begin{aligned}
B_3(s,Q,n) &\approx \sum_{k=1}^{n}\frac{I\!E((-(I_k^e+U))^+)^2 - I\!E((-(I_k^b+U))^+)^2}{2Q} \\
&\quad - \frac{I\!E\tilde{A}-I\!EA}{I\!EA}I\!ED\sum_{k=1}^{n}\frac{I\!E(-I_k^e)^+ - I\!E(-I_k^b)^+}{Q}.
\end{aligned} \tag{6.25}$$

Proof

For the derivation of the average backlog we will use the well-known relation that the inventory position equals the physical stock plus on order minus the backlog (see, for example, Hadley and Whitin (1963) page 187). The expected inventory position is equal to $s+Q/2$, and the expected amount on order is given by $\sum_{k=1}^{n}\frac{EDEL_{k:n}}{nEA}$. The latter can be shown analogously to the arguments of Hadley and Whitin. Imagine that orders flow into one end of a pipeline and procurements flow out of the other end. For $k \in \{1, \ldots, n\}$ the k-th partial delivery remains on average $\mathbb{E}L_{k:n}$ time units in the pipeline. A single demand unit has equal probability to be delivered from the k-th ($k \in \{1, \ldots, n\}$) partial delivery, and the expected flow out of the pipeline equals $\frac{\mathbb{E}D}{\mathbb{E}A}$. Therefore, the expected number of units in the pipeline should be $\sum_{k=1}^{n} \frac{\mathbb{E}D\mathbb{E}L_{k:n}}{n\mathbb{E}A}$. Hence,

$$B_3(s,Q,n) = B_4(s,Q,n) - (s + Q/2) + \sum_{k=1}^{n} \frac{\mathbb{E}D\mathbb{E}L_{k:n}}{n\mathbb{E}A}. \tag{6.26}$$

Note that for $k \in \{1, \ldots, n\}$ $\frac{\mathbb{E}D\mathbb{E}L_{k:n}}{\mathbb{E}A} = \mathbb{E}D(\sigma, \sigma + L_{k:n}) - \frac{\mathbb{E}\tilde{A} - \mathbb{E}A}{\mathbb{E}A}\mathbb{E}D$. Substitution of (6.24) into (6.26) yields

$$\begin{aligned}
B_3(s,Q,n) \approx{}& \frac{\mathbb{E}\tilde{A} - \mathbb{E}A}{\mathbb{E}A}\mathbb{E}D \sum_{k=1}^{n} \frac{\mathbb{E}(I_k^b)^+ - \mathbb{E}(I_k^e)^+}{Q} \\
&+ \sum_{k=1}^{n} \frac{\mathbb{E}((I_k^b + U)^+)^2 - \mathbb{E}((I_k^e + U)^+)^2}{2Q} - (s + Q/2) + \sum_{k=1}^{n} \frac{\mathbb{E}D\mathbb{E}L_{k:n}}{n\mathbb{E}A} \\
={}& \frac{\mathbb{E}\tilde{A} - \mathbb{E}A}{\mathbb{E}A}\mathbb{E}D\Big(\sum_{k=1}^{n} \frac{\mathbb{E}(I_k^b)^+ - \mathbb{E}(I_k^e)^+}{Q} - 1\Big) \\
&+ \sum_{k=1}^{n} \frac{\mathbb{E}((I_k^b + U)^+)^2 - \mathbb{E}((I_k^e + U)^+)^2}{2Q} \\
&- \frac{2sQ + Q^2}{2Q} + \sum_{k=1}^{n} \frac{\frac{2Q}{n}\mathbb{E}D(\sigma, \sigma + L_{k:n})}{2Q} \\
={}& -\frac{\mathbb{E}\tilde{A} - \mathbb{E}A}{\mathbb{E}A}\mathbb{E}D \sum_{k=1}^{n} \frac{\mathbb{E}(-I_k^e)^+ - \mathbb{E}(-I_k^b)^+}{Q} \\
&+ \sum_{k=1}^{n} \frac{\mathbb{E}(-(I_k^e + U)^+)^2 - \mathbb{E}(-(I_k^b + U)^+)^2}{2Q} \\
&+ \frac{1}{2Q}\Big[(s+Q)^2 - 2(s+Q)\mathbb{E}D(\sigma, \sigma + L_{n:n}) + \mathbb{E}D(\sigma, \sigma + L_{n:n})^2 \\
&\quad - s^2 + 2s\mathbb{E}D(\sigma, \sigma + L_{1:n}) - \mathbb{E}D(\sigma, \sigma + L_{1:n})^2 - 2sQ - Q^2 \\
&+ \sum_{k=2}^{n} \Big((s + \frac{k-1}{n}Q)^2 - 2(s + \frac{k-1}{n}Q)\mathbb{E}D(\sigma, \sigma + L_{k-1:n}) + \mathbb{E}D(\sigma, \sigma + L_{k-1:n})^2 \\
&\quad - (s + \frac{k-1}{n}Q)^2 + 2(s + \frac{k-1}{n}Q)\mathbb{E}D(\sigma, \sigma + L_{k:n}) - \mathbb{E}D(\sigma, \sigma + L_{k:n})^2\Big)\Big] \\
&+ \sum_{k=1}^{n} \frac{\frac{2Q}{n}\mathbb{E}D(\sigma, \sigma + L_{k:n})}{2Q} \\
={}& \sum_{k=1}^{n} \frac{\mathbb{E}((-(I_k^e + U))^+)^2 - \mathbb{E}((-(I_k^b + U))^+)^2}{2Q}
\end{aligned}$$

$$-\frac{\mathbb{E}\tilde{A}-\mathbb{E}A}{\mathbb{E}A}\mathbb{E}D\sum_{k=1}^{n}\frac{\mathbb{E}(-I_k^e)^+-\mathbb{E}(-I_k^b)^+}{Q}.$$

□

The objective is to minimize the sum of the holding, ordering, and backordering costs. The total relevant costs as function of the control parameters is given by the following expression

$$C(s,Q,n)=\frac{A(n)}{T(s,Q,n)}+b_3B_3(s,Q,n)+b_4B_4(s,Q,n)$$

Now we formulate the following minimization problem

$$\begin{aligned}(\mathcal{P}_1)\qquad &\text{minimize}\quad C(s,Q,n)\\ &\text{s.t.}\quad Q\geq 0;\\ &\qquad\quad n\in\mathbb{N}.\end{aligned}$$

When n is fixed, we can find the optimal values for s and Q, denoted by $s^*(n)$ and $Q^*(n)$ respectively, in the following way. For given values of n and Q, the optimal value of s can be determined by solving the equation $\frac{\partial C(s,Q,n)}{\partial s}=0$, presuming a unique solution exists. By using relations (6.19) and (6.27) it can be derived that

$$\frac{\partial C(s,Q,n)}{\partial s}=(b_3+b_4)P_3(s,Q,n)-b_3, \tag{6.27}$$

(see also section 2.3). Hence, for given values of n and Q the optimal value of s (denoted by $s^*(Q,n)$), can be determined by solving

$$P_3(s,Q,n)=\frac{b_3}{b_3+b_4}. \tag{6.28}$$

Since $P_3(s,Q,n)$ is increasing in s, and can take all values on (0,1), a unique solution indeed exists. Note the resemblance with the newsboy problem (see Silver and Peterson (1985, page 265)). So, for fixed n we can find $s^*(n)$ and $Q^*(n)$ by solving the following one-dimensional optimization problem

$$\begin{aligned}(\mathcal{P}_2)\qquad &\text{minimize}\quad C(s^*(Q,n),Q,n)\\ &\text{s.t.}\quad Q\geq 0.\end{aligned}$$

If we assume that $C(s^*(Q,n),Q,n)$ is convex in Q, we can determine $Q^*(n)$ by using for example Golden Section search, and $s^*(n)=s^*(Q^*(n),n)$. For practical situations we may restrict ourselves to a limited number of suppliers (n_{max}). For each n, $1\leq n\leq n_{max}$, we solve $(\mathcal{P}_2)$, and select that n for which the $C(s^*(n),Q^*(n),n)$ is minimal.

6.5.1 Validation of the algorithm

By simulation we first validate the proposed algorithm for computing the values of $B_3(s,Q,n)$, $B_4(s,Q,n)$, and $P_3(s,Q,n)$. The algorithm yields approximations for the values for the optimal decision parameters, because we assume that

Condition 6.4

(i) Replenishment orders do not cross

(ii) Exactly Q is ordered at a time.

(iii) $\tilde{A}$ is distributed as the forward recurrence time associated with F_A (see section 2.2).

(iv) U is distributed as the the forward recurrence time associated with F_D.

(v) The moments of $N(\sigma,\sigma+L_{k:n})$ are approximated by (2.17) and (2.18), which are asymptotic relations.

(vi) The distribution functions of $D(\sigma,\sigma+L_{k:n})+U$ and $D(\sigma,\sigma+L_{k:n})$ $(k=1,\ldots,n)$ are approximated by ME distributions;

(vii) $C(s^*(Q,n),Q,n)$ is convex in Q.

We will show that in spite of all these assumptions, our calculation scheme provides excellent approximations for the relevant performance characteristics given s, Q and n. Thereby the algorithm given at the end of the previous section yields near-optimal values for s^*, Q^* and n^*.

Note that assumptions (i), (ii) and (iii) are made for deriving approximations for $B_3(s,Q,n)$ and $B_4(s,Q,n)$. Assumptions (iv), (v), and (vi) are for computing the first two moments of $D(\sigma,\sigma+L_{k:n})$ and $D(\sigma,\sigma+L_{k:n})+U$ $(k=1,\ldots,n)$. While, assumption (vii) has to do with finding the optimal values for the decision variables.

It is well-known that for small values of Q as compared to $\mathbb{E}D(\sigma,\sigma+L_{n:n})$ assumption (i) is violated (see, for example, Kelle and Silver (1991b)). Assumptions (ii) and (iv) are violated only when Q is small with respect to $\mathbb{E}D$, i.e. $Q < Cond(D)$. Assumptions (iii) and (v) are violated when $\mathbb{E}L < Cond(A)$.

In practice assumption (i) may be violated when the number of suppliers is large. Therefore, we will investigate the effect of assumption (i) on the quality of the computation of the expected average physical stock level, expected average backlog level, and the fraction of the time the net stock is positive, by the proposed algorithm of section 6.2.

We used discrete event simulation to validate the quality of the approximations in terms of the deviation of the calculated performance measures by the algorithm described in section 6.2, and the performance measures computed by simulation. These experiments are done for a wide range of parameter values. The input values of the system parameters

Table 6.4: Basic parameter setting for the numerical experiments

n	$\mathbb{E}D$	c_D	$\mathbb{E}A$	c_A	$\mathbb{E}L$	c_L	Q	$P_{3,an}$
1	5	$\frac{1}{2}$	$\mathbb{E}D/5$	$\frac{1}{2}$	5	$\frac{3}{10}$	50	0.50
2	10	1		1	10	$\frac{1}{2}$	100	0.99
3		2		2		1	250	
5								
10								

are given in Table 6.4. For each of these 3240 experiments we calculated s by solving $P_3(s,Q,n) = P_{3,target}$ via a numerical search routine. $P_{3,target}$ represents the ratio $\frac{b_3}{b_3+b_4}$ and is chosen between two extreme value 0.50 ($b_3 = b_4$) and 0.99 ($b_3 = 99b_4$). The number of sub-runs performed in the simulation experiment is fixed to 5 (excluding the initialisation run), and the sub-run length is 100.000 time units. Furthermore, the demand sizes, interarrival times, and the lead times, are ME distributed. We computed $B_3(s,Q,n)$ and $B_4(s,Q,n)$ by formulas (6.19) and (6.27) which are denoted by $B_{3,an}$ and $B_{4,an}$ respectively. Simulation was used to verify whether $B_{3,an}$ and $B_{4,an}$ are equal to the related values computed by simulation, denoted by $B_{3,sim}$ and $B_{4,sim}$, respectively. We define $\Delta B_3 = |B_{3,an}$ - $B_{3,sim}|$ and $\Delta B_4 = |B_{4,an}$ - $B_{4,sim}|$.

Furthermore, we calculated by simulation the fraction of the partial deliveries that crossed any partial deliveries of previously placed replenishment orders, which is denoted by C_{sim}.

The results of these experiments are aggregated in Table 6.5, in which each line represents the average of the absolute deviations of the performance measures over 180 experiments. Since the mean absolute deviations of $B_{3,an}$ and $B_{4,an}$ have to be related to the absolute values of $B_{3,an}$ and $B_{4,an}$, we also computed the relative errors of the sum of inventory and backordering costs. That is, for $b_4 = 1$ and $b_3 = \{1, 10, 20\}$ we computed $\frac{|b_3(B_{3,an}-B_{3,sim})+b_4(B_{4,an}-B_{4,sim})|}{b_3 B_{3,sim}+b_4 B_{4,sim}}$ (see columns $b_3 = b_4$, $b_3 = 10b_4$, and $b_3 = 20b_4$).

From these experiments we can conclude the following about the quality of the expressions for the performance measures computed by the proposed algorithm in this section.

- For the situations in which $Q = 100$ or $Q = 250$, the algorithm performs good in all cases that are considered. Both the determination of s via $P_3(s,Q,n) = P_{3,an}$ and the computation of $B_3(s,Q,n)$ and $B_4(s,Q,n)$ yield accurate results.

- For the situations where $Q = 50, c_L = 1$, and $P_{3,an} = 0.50$, we see discrepancies between the target and achieved P_3-level. The explanation for this deviation is expressed by the fraction of partial deliveries that cross, which is in these situations up to 59 % of the partial deliveries.

- For high values of $P_{3,an}$ we note that $B_{3,an}$ deviates from $B_{3,sim}$. This has only a small impact on the computation of the sum of ordering and holding costs, which

Table 6.5: The deviations of simulation and the values calculated with the algorithm

Q	$P_{3,an}$	c_L	ΔB_4	ΔB_3	C_{sim}	$b_3 = b_4$	$b_3 = 10b_4$	$b_3 = 20b_4$
50	0.50	0.3	0.0532	0.0331	0.4889	0.0372	0.0320	0.0316
50	0.99	0.3	0.0014	0.5184	0.4892	0.0026	0.0139	0.0261
100	0.50	0.3	0.0248	0.0203	0.1446	0.0205	0.0194	0.0194
100	0.99	0.3	0.0012	0.3364	0.1447	0.0017	0.0082	0.0155
250	0.50	0.3	0.0061	0.0054	0.0048	0.0046	0.0051	0.0052
250	0.99	0.3	0.0012	0.2022	0.0047	0.0013	0.0035	0.0064
50	0.50	0.5	0.0804	0.0480	0.5376	0.0616	0.0510	0.0501
50	0.99	0.5	0.0022	0.7672	0.5374	0.0042	0.0222	0.0421
100	0.50	0.5	0.0224	0.0214	0.1987	0.0201	0.0203	0.0204
100	0.99	0.5	0.0015	0.5224	0.1987	0.0026	0.0142	0.0269
250	0.50	0.5	0.0065	0.0064	0.0085	0.0051	0.0059	0.0061
250	0.99	0.5	0.0012	1.1435	0.0085	0.0014	0.0049	0.0091
50	0.50	1.0	0.2627	0.0973	0.5878	0.1662	0.1154	0.1120
50	0.99	1.0	0.0040	0.9804	0.5878	0.0070	0.0338	0.0637
100	0.50	1.0	0.0801	0.0608	0.3094	0.0703	0.0628	0.0623
100	0.99	1.0	0.0039	0.9254	0.3091	0.0068	0.0333	0.0626
250	0.50	1.0	0.0075	0.0157	0.0377	0.0095	0.0137	0.0143
250	0.99	1.0	0.0031	0.4538	0.0377	0.0037	0.0181	0.0344

follows from the last columns in Table 6.5. This can be explained by the fact that for large values of b_3, the determination of the optimal values for the decision variables is basically a trade-off between the ordering and holding costs.

- Interestingly, the crossing of orders does not influence the quality of the approximations for high values of $P_{3,an}$, that is, high values of b_3.

These results point out that the proposed algorithm performs very well. We have to be careful only in situations where crossing of orders frequently occurs, or cases with low values of b_3. The probability of order crossing can be determined by computing

$$\mathbb{P}(L_{n:n} > T + L_{1:n}) \tag{6.29}$$

where T represents the length of a replenishment cycle. In Chapter 7 we present formulas to determine the first two moments of T for an (s, Q) inventory model. It can be argued that these formulas are also applicable for the multiple sourcing situation. So, in case the distribution functions of $L_{1:n}$, $L_{n:n}$ and T are approximated by ME-distributions, the crossing probability (6.29) can be computed easily.

In the following experiment we checked numerically whether assumption (vii) is valid, i.e. if $C(s^*(Q,n), Q, n)$ is convex in Q. Of course this is not the appropriate way of validating the convexity assumption. However, we have not been able to derive tractable conditions for convexity. Therefore, we resort to a numerical investigation into the convexity of $C(s^*(Q,n), Q, n)$. For these experiments we fixed the following input values, $(\mathbb{E}D, c_D) = (10, 1)$, $(\mathbb{E}A, c_A) = (1, 1)$, $\mathbb{E}L = 10$, and $b_4 = 0.01$. In Figures 6.2 to 6.5 we plotted $C(s^*(Q,n), Q, n)$ as function of Q. We did not find any numerical counter examples of the conjecture that $C(s^*(Q,n), Q, n)$ is convex. These figures show also that $Q^*(n)$ is increasing in n. Moreover, the optimal number of suppliers is depending on the input parameters. The cost parameters K, c, b_3, and b_4 indeed influence n^* (compare Figures 6.2, 6.4 and 6.5). But also the parameters of the underlying lead time process do influence n^* (compare Figure 6.2 with Figure 6.3).

6.5.2 The optimal number of suppliers

From Figures 6.2 to 6.5, it is clear that n^* depends on the values of the input parameters. Therefore, we designed a number of experiments to get some insight into the optimal number of suppliers.

First of all we compared our results with the results presented by Ramasesh et al.(1991). Ramasesh et al. consider the same objective, under constant demand and with at most two suppliers. Hence the model is a special case of the model discussed in this thesis. We fit the parameters of our model to the parameters of the model in Ramasesh et al.(1991) as follows. By considering small interarrival time of customers and low coefficient of variations of D and A, we can approximate the model considered by Ramasesh et al.(1991). Moreover, the ordering costs in Ramasesh et al. (1991) for the two supplier situation are given by

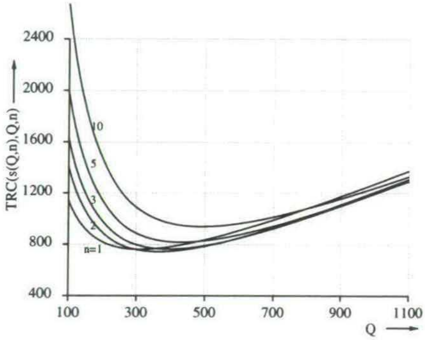


Figure 6.2: $C(s(Q,n),Q,n)$ as function of Q, where $c_L = 0.3$, $K = 30$, $c = 0.5$, and $b_3 = 0.1$.

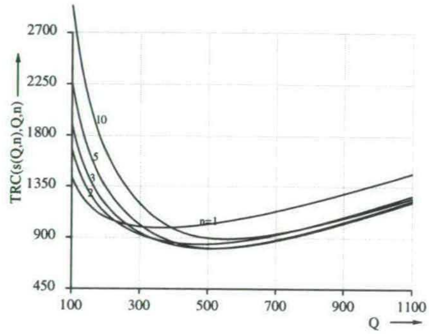


Figure 6.3: $C(s(Q,n),Q,n)$ as function of Q, where $c_L = 1$, $K = 30$, $c = 0.5$, and $b_3 = 0.1$.

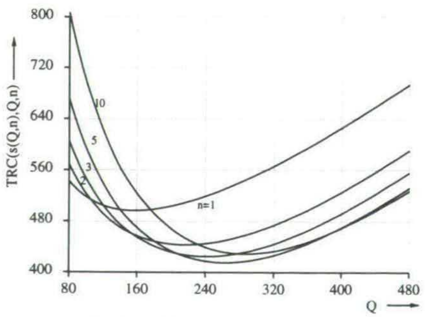


Figure 6.4: $C(s(Q,n),Q,n)$ as function of Q, where $c_L = 0.5$, $K = 5$, $c = 0.5$, and $b_3 = 0.1$.

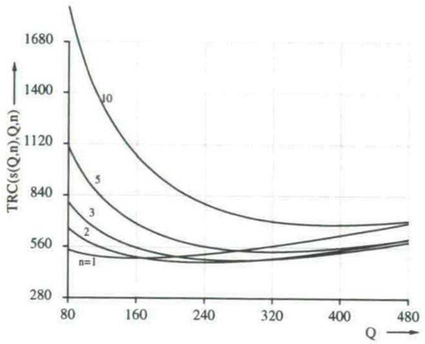


Figure 6.5: $C(s(Q,n),Q,n)$ as function of Q, where $c_L = 0.5$, $K = 5$, $c = 1$, and $b_3 = 0.1$.

Table 6.6: Comparison results from Ramasesh et al. (1991) with our results

		$n=1$			$n=2$			$n=3$				
c		s_1^*	Q_1^*	C^*	s_2^*	Q_2^*	C^*	s_3^*	Q_3^*	C^*	n^*	C^*
0.00	(J)	189	1275	1222	34	1334	978	-7	1309	890	∞	-
	(R)	191	1271	1220	36	1333	981					
0.263	(J)	189	1275	1222	23	1407	1051	-22	1418	1013	4	1010
	(R)	191	1271	1220	24	1408	1054					
0.678	(J)	189	1275	1222	5	1539	1186	-52	1651	1264	2	1186
	(R)	191	1271	1220	6	1542	1190					
0.761	(J)	189	1275	1222	2	1569	1217	-59	1708	1324	2	1217
	(R)	191	1271	1220	2	1573	1220					

$A(2) = \alpha K$, where $\alpha \in [1,2]$. Hence, the appropriate choice of c is $\log_2 \alpha$. The results for $n = 1$ and $n = 2$ are similar, see Table 6.6, where (J) denotes the results of our model and (R) the results of the model of Ramasesh et al. (1991). The optimal values of s and Q and the value of the total relevant costs are almost equal. For the situation that $\alpha = 1$ (i.e. $c = 0$) the optimal number of suppliers is infinity, as was already noted by Larson(1989). Furthermore, we note that for values of $\alpha > 1$, using two suppliers can be advantageously, but is not optimal (see $c = 0.263$).

In the experiments that follow we take one day as the basic time unit, and one year equal to 250 (working) days. We investigate the effect of the cost parameters K, c, b_3 and b_4 on the optimal number of suppliers. We fixed the following values for the system parameters: $(\mathbb{E}D, c_D) = (10, 1)$, $(\mathbb{E}A, c_A) = (1, 1)$, and $(\mathbb{E}L, c_L) = (10, 0.5)$. We fixed c equal to 0.5 and the inventory holding cost b_4 equal to 0.04. This represents an article with purchase price of \$40 and a opportunity factor of 0.25 \$/\$/year. First we varied b_3 between 1, 10, 100, and 1000 times b_4, and for each setting we calculated the optimal number of suppliers as function of K (see Figures 6.6 and 6.7). The number of values chosen for K is equal to 100 for each value of b_3. To generate Figures 6.6 and 6.7 required about 14 minutes CPU time on a SUNSPARC-station 4. We see that $n^* \to \infty$ when $K \downarrow 0$, and $n^* = 1$ when $K \to \infty$, which is also intuitively clear. Moreover, n^* increases when $\frac{b_3}{b_3+b_4}$ increases. And n^* decreases when c increases (compare Figures 6.6 with 6.7), which is intuitively clear, as well.

In the final experiments we investigate the effect of the parameters of the underlying stochastic processes $(\mathbb{E}D, c_D)$, $(\mathbb{E}A, c_A)$, and $(\mathbb{E}L, c_L)$ on n^*. We considered situations in which $K = 20$, $c = 0.5$, $h = 0.04$, and $b = 0.4$. We started with $(\mathbb{E}D, c_D) = (10, 1)$, $(\mathbb{E}A, c_A) = (1, 1)$, and $(\mathbb{E}L, c_L) = (10, 0.5)$, as in the basic situation, however in each experiment we varied one or two of these system parameters.

In Figure 6.8 we computed n^* as function of $\mathbb{E}D$, for various values of $\mathbb{E}L$. We note that n^* is almost linear in both $\mathbb{E}D$ and $\mathbb{E}L$. In Figure 6.9 we varied $\mathbb{E}A$. Similar to the effect of K, we see that $n^* \to \infty$ when $\mathbb{E}A \downarrow 0$, and $n^* = 1$ when $\mathbb{E}A \to \infty$.

In case the coefficients of variation of D and A are varied, we only find minor effects on

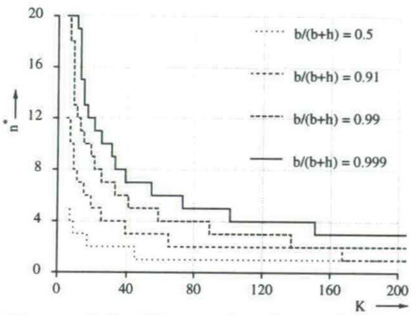


Figure 6.6: The optimal number of suppliers as function of K with $c = 0.5$.

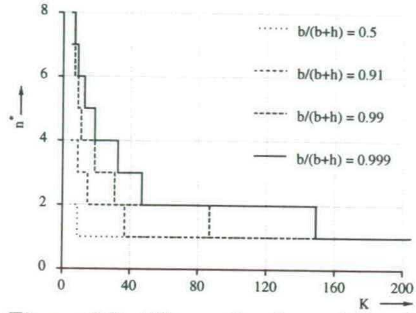


Figure 6.7: The optimal number of suppliers as function of K with $c = 1$.

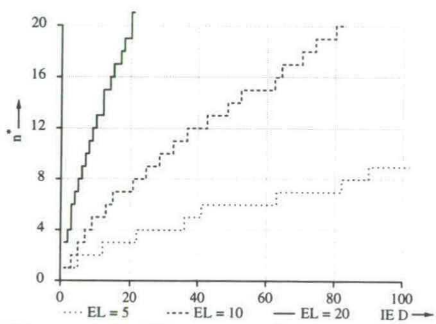


Figure 6.8: The optimal number of suppliers as function of $\mathbb{E}D$.

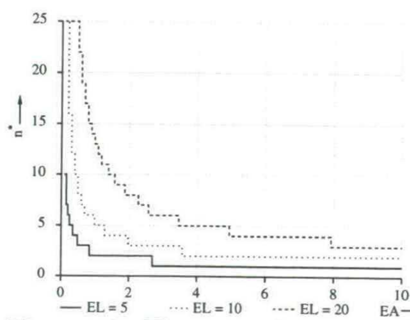


Figure 6.9: The optimal number of suppliers as function of $\mathbb{E}A$.

the optimal number of suppliers. In Figure 6.10 we varied c_A. It is important to note that higher values of c_A can lead to both lower and higher values of n^*. A detailed investigation of the solutions is given for $\mathbb{E}L = 20$ and c_A is varied between 1.1 and 1.2 (see Table 6.7). The differences between $C(s^*(12), Q^*(12), 12)$ and $C(s^*(13), Q^*(13), 13)$ are very small, and for some values of c_A the $C(s^*(12), Q^*(12), 12)$ is smaller than $C(s^*(13), Q^*(13), 13)$ and for other values the other way around. When n increases, the optimal reorder point will decrease, however, the optimal reorder quantity will increase. Hence, the inventory holding costs may increase or decrease.

The impact of c_D on n^* are similar to the effects of c_A. In contrast with this, n^* turns out to be very sensitive to the value of c_L. In Figure 6.11 we varied both $\mathbb{E}L$ and c_L. This sensitivity can be explained by considering effects of c_L that interfere. When c_L increases, the first orders will arrive earlier, which leads to lower values of the reorder point. But due to the earlier arrival of the partial deliveries the expected average physical stock will slightly increase. Finally, it is noteworthy that often there are only minor differences in

Table 6.7: Detailed investigation of the solutions

c_A	n	$C(s^*(n), Q^*(n), n)$	$s^*(n)$	$Q^*(n)$
1.1	1	3051.58	290.78	184.21
	12	2091.28	95.36	445.31
	13	2090.65	93.26	447.99
	14	2091.41	91.41	450.39
1.2	1	3091.85	291.86	185.81
	12	2146.90	97.01	452.29
	13	2146.91	94.88	454.77
	14	2147.94	91.35	457.08

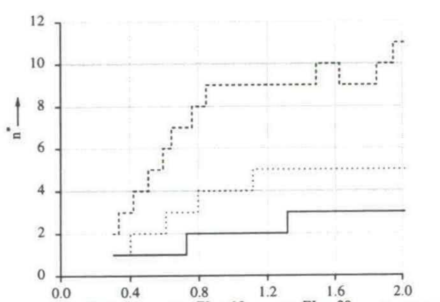


Figure 6.10: The optimal number of suppliers as function of c_A.

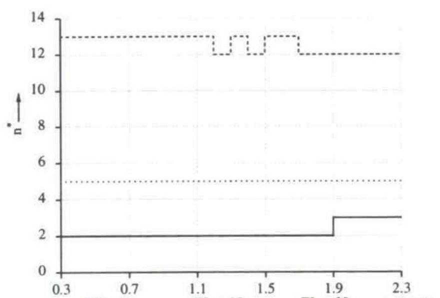


Figure 6.11: The optimal number of suppliers as function of c_L.

the total relevant cost for two successive values of n (see, for example Table 6.7).

6.6 Conclusions and future research

In this chapter an (s, Q) inventory model is presented with order splitting, where the demand is modelled as a compound renewal process, and lead times of the suppliers are independent and identically distributed random variables. This model can be applied to many practical situations. We derived different expressions for the P_1 service measures. We discussed that the non-stock out probability (P_1 service measure) in an order splitting environment is not appropriate.

When shortage cost are hard to specify, a service level constraint can be used to determine the reorder point s. Based only on the first two moments of the underlying demand and lead time process, an algorithm is derived to compute s by solving $P_i(s, Q, n) = P_{i,target}$ for $i = 2, 3$. The algorithm turns out to perform very good for situations in which the number of order crossings was not too high.

We derived expressions for the expected average physical stock, the expected average backlog level, and the fraction of the time that the physical stock is positive. Furthermore, an algorithm is derived to compute these performance measures based only on moments of the underlying demand and lead time process. The algorithm turned out to perform very good for situations in which the number of order crossings was not too high. Although the performance measures are also valid for non-identically distributed lead times of suppliers, the algorithm is only developed for identically distributed lead times. In Balakrishnan (1988) recurrence relations for order statistics from n independent and non-identically distributed random variables are given. However, the actual computation of the moments of the order statistics is quite cumbersome. Clearly this is a topic of future research.

We considered the problem of determining the appropriate values for the control parameters s, Q, and n. We minimized the sum of ordering, holding, and backordering costs. The optimal number of suppliers turned out to be very sensitive for the combination of input parameters. A striking observation was that n^* is not always increasing when the coefficient of variation of the lead times does. The algorithm can be used to generate graphical support instantaneously for a wide range of input values.

In this chapter we do not consider the optimization of the fractions of the purchase volume of size Q which is purchased at the different suppliers, as is done in Lau and Lau (1994) and in Chapter 5. The extension in which $Q_1, \ldots, Q_n$ not necessarily of equal size is straightforward with respect to deriving the expressions for the performance measures. However, the numerical computation of these performance measures is quite cumbersome. The reason for this is that the allocation of a specific fraction of the purchase volume to a specific supplier makes it necessary to condition on the sequence in which the suppliers deliver. In case of n suppliers $n!$ sequences are possible, which indicates the computational complexity of this situation.

Appendix 6.A: The derivation of the moments of $L_{k:n}$

Balakrishnan and Cohen (1991) presented a method for computing the moments of order statistics in case the common distribution function is from the class of Gamma distributions. We extended this method for the case of ME-distribtions. For fitting a ME-distribution Tijms (1994) distinguishes between the situations that coefficient of variation is less than one and larger than one. In the first case a mixture of a Erlang $k-1$ and Erlang k distribution with the same scale parameter is suggested, see section 2.4. Hence the density function is given by

$$g(x) = p\mu^{k-1}\frac{x^{k-2}}{(k-2)!}e^{-\mu x} + (1-p)\mu^k\frac{x^{k-1}}{(k-1)!}e^{-\mu x} \tag{6.30}$$

and

$$1 - G(x) = \sum_{j=0}^{k-2}\frac{(\mu x)^j}{j!}e^{-\mu x} + (1-p)\frac{(\mu x)^{k-1}}{(k-1)!}e^{-\mu x} \tag{6.31}$$

From (6.31) we see that $((1-G(x))e^{\mu x})^n$ is a polynoom of order $(k-1)n$ and we define it coefficients as $b_j(n)$ for $j = 0, \ldots, (k-1)n$. The coefficient $b_j(n)$ can easily be computed recursively from

$$b_j(n) = \sum_{i=min\{0,j-(n-1)(k-1)\}}^{max\{k-1,j\}} b_i(1)b_{j-i}(n-1) \quad j = 0, \ldots, (k-1)n. \tag{6.32}$$

Now

$$\begin{aligned}
I\!\!EL_{1:n}^m &= n\int_0^\infty x^m(1-G(x))^{n-1}dG(x) \\
&= \frac{n}{\mu^m}\int_0^\infty (1-G(x))^{n-1}(\mu x)^m\mu\Big(p\frac{(\mu x)^{k-2}}{(k-2)!} + (1-p)\frac{(\mu x)^{k-1}}{(k-1)!}\Big)e^{-\mu nx}dx \\
&= \frac{n}{\mu^m}\sum_{j=0}^{(k-1)(n-1)} b_j(n-1)\int_0^\infty \mu\Big(p\frac{(\mu x)^{j+m+k-2}}{(k-2)!} + (1-p)\frac{(\mu x)^{j+m+k-1}}{(k-1)!}\Big)e^{-\mu nx}dx \\
&= \frac{n}{\mu^m}\Bigg(\sum_{j=0}^{(k-1)(n-1)} b_j(n-1)\Big(\frac{p(j+m+k-2)!}{n^{j+m+k-1}(k-2)!}\int_0^\infty (\mu n)^{j+m+k-1}\frac{x^{j+m+k-2}}{(j+m+k-2)!}e^{-\mu nx}dx\Big) \\
&\quad + \sum_{j=0}^{(k-1)(n-1)} b_j(n-1)\Big(\frac{(1-p)(j+m+k-1)!}{n^{j+m+k}(k-1)!}\int_0^\infty (\mu n)^{j+m+k}\frac{x^{j+m+k-1}}{(j+m+k-1)!}e^{-\mu nx}dx\Big)\Bigg) \\
&= \frac{n}{\mu^m}\sum_{j=0}^{(k-1)(n-1)} b_j(n-1)\Big(\frac{p(j+m+k-2)!}{n^{j+m+k-1}(k-2)!} + \frac{(1-p)(j+m+k-1)!}{n^{j+m+k}(k-1)!}\Big).
\end{aligned}$$

Using relation 3.3.3. from Balakrishnan (1991) we then obtain $\mathbb{E}L^m_{i:n}$ for $i = 2, \ldots, n$

$$\mathbb{E}L^m_{i:n} = \sum_{j=n-i+1}^{n} (-1)^{j-n+i-1} \binom{n}{j} \binom{j-1}{n-i} \mathbb{E}L^m_{1:j} \tag{6.33}$$

For the situation that coefficient of variation is larger than one Tijms (1994) suggests a Hyper-exponential distribution. Hence the density function is given by

$$g(x) = p\mu_1 e^{-\mu_1 x} + (1-p)\mu_2 e^{-\mu_2 x} \tag{6.34}$$

and

$$1 - G(x) = pe^{-\mu_1 x} + (1-p)e^{-\mu_2 x} \tag{6.35}$$

Let $\lambda_1(j,n) = \mu_1(j+1) + \mu_2(n-1-j)$ and $\lambda_2(j,n) = \mu_1 j + \mu_2(n-j)$, then

$$\begin{aligned}
\mathbb{E}L^m_{1:n} &= n\int_0^\infty x^m(1-G(x))^{n-1}dG(x) \\
&= n\int_0^\infty x^m(pe^{-\mu_1 x} + (1-p)e^{-\mu_2 x})^{n-1}(p\mu_1 e^{-\mu_1 x} + (1-p)\mu_2 e^{-\mu_2 x})dx \\
&= n\mu_1 p \sum_{j=0}^{n-1} \binom{n-1}{j} \int_0^\infty x^m p^j (1-p)^{n-1-j} e^{-\mu_1(j+1)x} e^{-\mu_2(n-1-j)x} dx \\
&\quad + n\mu_2(1-p) \sum_{j=0}^{n-1} \binom{n-1}{j} \int_0^\infty x^m p^j (1-p)^{n-1-j} e^{-\mu_1 j x} e^{-\mu_2(n-j)x} dx \\
&= n\mu_1 p \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \frac{m!}{\lambda_1(j,n)^{m+1}} \int_0^\infty \lambda_1(j,n)^{m+1} \frac{x^m}{m!} e^{\lambda_1(j,n)x} dx \\
&\quad + n\mu_2(1-p) \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \frac{m!}{\lambda_2(j,n)^{m+1}} \int_0^\infty \lambda_2(j,n)^{m+1} \frac{x^m}{m!} e^{\lambda_2(j,n)x} dx \\
&= \mu_1 np(1-p)^{n-1} m! \sum_{j=0}^{n-1} \binom{n-1}{j} (\frac{p}{1-p})^j \lambda_1(j,n)^{-(m+1)} \\
&\quad + \mu_2 n(1-p)^n m! \sum_{j=0}^{n-1} \binom{n-1}{j} (\frac{p}{1-p})^j \lambda_2(j,n)^{-(m+1)}
\end{aligned}$$

Again we can use relation (6.33) to compute $\mathbb{E}L^m_{i:n}$ for $i = 2, \ldots, n$.

Chapter 7

Large order overflow

In general, a stockpoint in a multi-echelon distribution chain satisfies the demand of all customers that arrive at that stockpoint, where customers are defined as the external customers as well as replenishment orders of downstream stockpoints in the distribution chain. However, in case large order overflow is applied, customers with large demand are not satisfied by the stockpoint at which they arrive, but by an alternative (e.g. an upstream) stockpoint. Under large order overflow for each stockpoint i a maximal customer order quantity $D_{max,i}$ and an alternative stockpoint r_i are given, such that customers with demand larger than $D_{max,i}$ are satisfied by the source r_i instead of stockpoint i itself. Re-routing orders to an upstream source implies a decrease of the number of internal replenishments, which means lower ordering costs of the downstream stockpoints. Since the total external demand does not change, the ordering costs of the upstream stockpoints remain the same. Furthermore, safety stocks of the downstream stockpoints decrease when large order overflow is applied. Consequently, the holding costs at the downstream stockpoints decrease. On the other hand, lead times or transportation costs for re-routed customers will increase. Therefore it may not be easy to persuade customers to accept large order overflow and one may need to give discounts or one must provide fast transportation to obtain the customers' willingness to collaborate.

This chapter provides an heuristic algorithm to quantify the savings in ordering and holding costs and the increase in (emergency) transportation costs due to large order overflow. This algorithm enables management to make a proper trade-off between cost savings and cost increases. When the optimal value for $D_{max,i}$ (denoted by $D^*_{max,i}$) approaches infinity, large order overflow is not cost effective. On the other hand, when $D^*_{max,i}$ approaches zero, almost all customers are re-routed to the alternative stockpoint, i.e. the stockpoint becomes obsolete. Hence, based on the algorithm, not only management is able to decide whether or not to apply large order overflow, but it also can decide to reduce the number of stockpoints in the distribution chain.

Alternatively, one could consider small order underflow. In case small order underflow is applied, customers with small demand are not satisfied by the stockpoint at which they arrive, but by an alternative stockpoint (e.g. an downstream stockpoint). This chapter only deals with large order overflow.

First we will consider a serial system with two stockpoints to provide some insights into the profitability of large order overflow. Secondly, we will investigate a two-echelon system with N local stockpoints to provide some indications about the order of magnitude of the savings that can be obtained in real-world situations.

7.1 The serial network

Consider a serial network with two stockpoints. Stockpoint 1 represents a manufacturer or a power-retailer, whereas stockpoint 2 represents a retailer supplying a local market. Stockpoint 1 supplies stockpoint 2. Both stockpoints i ($i = 1, 2$) face independent compound Poisson demand processes with interarrival times A_i and demand sizes D_i. Only at stockpoint 2 large order overflow is applied. Customers arriving at stockpoint 2 with demand larger than D_{max} are re-routed to stockpoint 1. The replenishments at stockpoint i ($i = 1, 2$) are controlled by an (s_i, Q_i) policy. The inter replenishment times and replenishment sizes initiated at stockpoint i are denoted by (T_i, O_i).

The deliveries from stockpoint i ($i = 1, 2$) are described by the demand process (A_i^d, D_i^d) whereas the re-routed customers of stockpoint i are described by the demand process (A_i^r, D_i^r). Using that (A_2, D_2) is a compound Poisson process and that splitting or merging of a Poisson process yields again one or more Poisson processes (see, e.g., Theorem 1.2.3. Tijms (1994)), we see that (A_2^d, D_2^d) and (A_2^r, D_2^r) are independent compound Poisson processes. When (A_2, D_2) is a compound renewal process, (A_2^d, D_2^d) and (A_2^r, D_2^r) are dependent processes. In Figure 7.1 we have depicted the actual material flow in terms of various demand and replenishment processes. For example, the deliveries out of stockpoint 1 (A_1^d, D_1^d) consists of the direct external demand process (A_1, D_1), the replenishment orders for stockpoint 2 (T_2, O_2) and the overflow demand stream from stockpoint 2 (A_2^r, D_2^r).

The transportation time to stockpoint i is a random variable L_i. Demand which cannot be delivered directly from shelf is backordered. The waiting time of an arbitrary customer arriving at stockpoint i is denoted by W_i. Hence, the actual lead time of a replenishment order of stockpoint 2 is equal to the transportation time to stockpoint 2 (L_2) plus the waiting time at stockpoint 1 (W_1).

The problem is to find values for the control variables s_i, Q_i ($i = 1, 2$) and D_{max}. The objective is to minimize the sum of ordering, holding and transportation costs subject to a service level constraint at each stockpoint. As service measure we use the P_2 service measure, see section 2.1.4, where $P_{2,i}$ denotes the P_2 service level at stockpoint i.

Ordering costs are fixed per replenishment and are denoted by a_i at stockpoint i. The transportation from stockpoint i to external customers are fixed per transportation, i.e. independent of the size, and are denoted by k_i. We assume that the transportation costs of replenishments to stockpoint i are included in a_i. Furthermore, we assume that $k_1 > k_2$ which justifies the use of two stockpoints. The inventory holding costs are proportional to the average physical stock with rate b_4 at both stockpoints. Assuming constant inventory holding costs along the supply chain is based on the fact that once a product has been completed the material value remains constant during transportation through the supply

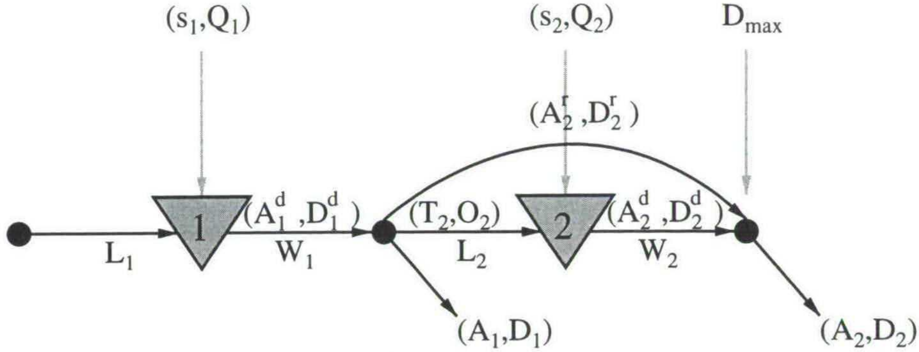


Figure 7.1: The serial network

chain. Let $B_{4,i}$ be the average physical stock at stockpoint i. Then the total relevant costs as function of the control parameters is given by the following expression

$$
\begin{aligned}
C(s_1, Q_1, s_2, Q_2, D_{max}) &= \sum_{i=1}^{2}\left(\frac{a_i}{\mathbb{E}T_i} + b_4 B_{4_i}\right) + \frac{k_1}{\mathbb{E}A_1} + \frac{k_1}{\mathbb{E}A_2^r} + \frac{k_2}{\mathbb{E}A_2^d} \\
&= \sum_{i=1}^{2}\left(\frac{a_i}{\mathbb{E}T_i} + b_4 B_{4_i} + \frac{k_i}{\mathbb{E}A_i}\right) + \frac{k_1 - k_2}{\mathbb{E}A_2^r} \qquad (7.1)
\end{aligned}
$$

Now we formulate the following minimization problem

$$
\begin{aligned}
(\mathcal{P}_1) \quad & \text{minimize} \quad C(s_1, Q_1, s_2, Q_2, D_{max}) \\
& \text{s.t.} \quad P_{2,1}(s_1, Q_1, s_2, Q_2, D_{max}) = P_{2,1,target}; \\
& \qquad\;\; P_{2,2}(s_1, Q_1, s_2, Q_2, D_{max}) = P_{2,2,target}.
\end{aligned}
$$

7.1.1 The solution procedure

In this section a heuristic method is described for solving $(\mathcal{P}_1)$. The outline of the heuristic is as follows. We solve $(\mathcal{P}_1)$ for fixed D_{max} and then use local search to find the optimal value of D_{max}. The heuristic to solve $(\mathcal{P}_1)$ for fixed D_{max} is based on a decomposition of the network into two single echelon (s, Q) inventory systems. Her we follow de Kok (1996). Therefore, we may use formulas (2.27) and (2.33) for $P_{2,i}$ and $B_{4,i}$ $i = 1, 2$ respectively. Let $D_i(0, L_i)$ be the demand during the lead time at stockpoint i. Then

$$\begin{aligned} P_{2,i}(s_i, Q_i) &= 1 - \frac{\mathbb{E}(D_i(0,L_i) + U_i - s_i)^+ - \mathbb{E}(D_i(0,L_i) + U_i - s_i - Q_i)^+}{Q_i}; \qquad (7.2) \\ B_{4,i}(s_i, Q_i) &= \frac{\mathbb{E}\tilde{A}_i^d - \mathbb{E}A_i^d}{\mathbb{E}A_i^d} \mathbb{E}D_i^d \frac{\mathbb{E}(s_i + Q_i - D_i(0,L_i) - U_i)^+}{Q_i} \\ &\quad - \frac{\mathbb{E}\tilde{A}_i^d - \mathbb{E}A_i^d}{\mathbb{E}A_i^d} \mathbb{E}D_i^d \frac{\mathbb{E}(s_i - D_i(0,L_i) - U_i)^+}{Q_i} \\ &\quad + \frac{\mathbb{E}((s_i + Q_i - D_i(0,L_i))^+)^2 - \mathbb{E}((s_i - D(_i0,L_i))^+)^2}{2Q_i}, \qquad (7.3) \end{aligned}$$

where $\tilde{A}_i^d$ be the asymptotic forward recurrence time associated with A_i^d.
For a given value of D_{max} we can determine (A_i^d, D_i^d) at both stock points, where (A_1^d, D_1^d) is a superposition of three demand processes. For both stockpoints we approximate the optimal values for Q_i by the economic order quantity, denoted by $Q_i^*(D_{max})$. Since $P_{2,1}$ is independent of s_2 we can determine the optimal reorder point for stockpoint 1, denoted by $s_1^*(D_{max})$. This is done by solving s_1 from the service equation. Now we can determine the customer waiting times at stockpoint 1. Combining W_1 and L_2 enables us to determine the actual lead time at stockpoint 2 and from that the optimal reorder point $s_2^*(D_{max})$ by solving s_2 from the service equation.
By assuming that $C(s_1^*(D_{max}), Q_1^*(D_{max}), s_2^*(D_{max}), Q_2^*(D_{max}), D_{max})$ is convex in D_{max}, and using local search, enables us to find the optimal value for D_{max}.

Before going into the details of each step we summarize below the heuristic for solving $s_1^*(D_{max}), Q_1^*(D_{max}), s_2^*(D_{max}), Q_2^*(D_{max})$ for a given value of D_{max} (in the sequel we will suppress the (D_{max}) in $s_1^*(D_{max}), Q_1^*(D_{max}), s_2^*(D_{max}), Q_2^*(D_{max})$) .

(1) Using (A_2, D_2) and D_{max} determine (A_2^d, D_2^d) and (A_2^r, D_2^r).

(2) With (A_1, D_1), (A_2, D_2) and (A_2^d, D_2^d) determine Q_1^* and Q_2^*.

(3) Given (A_2^d, D_2^d) and Q_2^* determine (T_2, O_2).

(4) Using (A_1, D_1), (A_2^r, D_2^r) and (T_2, O_2) determine (A_1^d, D_1^d).

(5) Given (A_1^d, D_1^d), L_1, Q_1^* and $P_{2,1,target}$ determine s_1^* by solving s_1 from the service level constraint.

(6) Using (A_1^d, D_1^d), L_1, Q_1^* and s_1 determine W_1.

(7) Using (A_2^d, D_2^d), L_2, W_1, Q_2^* and $P_{2,2,target}$ determine s_2^* by solving s_2 from the service level constraint.

ad (1) The large order overflow rule splits the original compound Poisson demand process at stockpoint 2 into two independent compound Poisson demand processes. It is easy to see that

$$\mathbb{E}A_2^d = \frac{\mathbb{E}A_2}{F_{D_2}(D_{max})}, \tag{7.4}$$

$$\mathbb{E}(D_2^d)^k = \frac{\int_0^{D_{max}} x^k dF_{D_2}(x)}{F_{D_2}(D_{max})}, \tag{7.5}$$

and

$$\mathbb{E}A_2^r = \frac{\mathbb{E}A_2}{1 - F_{D_2}(D_{max})}, \tag{7.6}$$

$$\mathbb{E}(D_2^r)^k = \frac{\int_{D_{max}}^{\infty} x^k dF_{D_2}(x)}{1 - F_{D_2}(D_{max})}. \tag{7.7}$$

ad (2) Approximations for the optimal values of Q_i^* $(i = 1, 2)$ are given by the economic order quantities. It has been shown that the Economic Order Quantity is quite robust (see, e.g., section 3.3). Hence,

$$Q_1^* = \sqrt{\frac{2a_1(\frac{ED_1}{EA_1} + \frac{ED_2}{EA_2})}{b_4}}, \tag{7.8}$$

$$Q_2^* = \sqrt{\frac{2a_2\frac{ED_2^d}{EA_2^d}}{b_4}}. \tag{7.9}$$

ad (3) The first two moments of the interarrival times and sizes of replenishment orders are independent of the reorder point. We confine ourselves to a brief exposition of the determination of (T_2, O_2) (for a detailed analysis we refer to de Kok, Pyke and Baganha (1996)). The size of a replenishment order (i.e. the number of multiples of Q_2^* that are ordered simultaneously) depends on the undershoot of the reorder level. The first two moments of O_2 are given by the following expressions (see de Kok, Pyke and

Baganha 1996))

$$\mathbb{E}O_2 = \frac{Q_2^* \mathbb{E}D_2^d}{\int_0^{Q_2^*}(1 - F_{D_2^d}(x))dx}, \tag{7.10}$$

$$\mathbb{E}O_2^2 = (Q_2^*)^2 \sum_{k=0}^{\infty}(2k+1)(1 - F_{U_2}(kQ_2^*)), \tag{7.11}$$

where U_2 denotes the undershoot under s_2, the c.d.f. of which is given by

$$F_{U_2}(u) = 1 - \frac{\int_0^{Q_2^*}(1 - F_{D_2^d}(u+x))dx}{\int_0^{Q_2^*}(1 - F_{D_2^d}(x))dx}. \tag{7.12}$$

The determination of an exact expression for the first two moments of the inter-replenishment times is in general intractable. We resort to an approximate scheme. The inter-replenishment time is equal to the total time that is needed to let arrive enough customers so that their total demand is larger than Q_2^*. Hence the appropriate formulas for the moments of T_2 are equal to (2.14) and (2.15)

$$\mathbb{E}T_2 = \mathbb{E}N(0, Q_2^*)\, \mathbb{E}A_2^d, \tag{7.13}$$

$$\mathbb{E}T_2^2 = \mathbb{E}N(0, Q_2^*)\, \sigma^2(A_2^d) + \mathbb{E}N(0, Q_2^*)^2(\mathbb{E}A_2^d)^2, \tag{7.14}$$

where $N(0, Q_2^*)$ denotes the number of customers that is required to exceed a total demand of Q_2^*, i.e. the number of customers arriving during a replenishment cycle. We approximate the first two moments of $N(0, Q_2^*)$ by its asymptotic moments. The inventory position just after an order epoch equals $s_2^* + nQ_2^* - U_2$, where n is such that $s_2^* < s_2^* + nQ_2^* - U_2 \leq s_2^* + Q_2^*$. Then the demand sizes form a delay renewal process with respect to the inventory position. To compute the first two moments of $N(0, Q_2^*)$ we see that the time stationary situation applies. Therefore we can use (2.20) and (2.21). Let $\alpha_i^j := (\mathbb{E}(D_2^d)^i)^j$. Then

$$\mathbb{E}N(0, Q_2^*) \approx \frac{Q_2^*}{\alpha_1^1}, \tag{7.15}$$

$$\mathbb{E}N(0, Q_2^*)^2 \approx \frac{(Q_2^*)^2}{\alpha_1^2} + Q_2^*\left(\frac{\alpha_2^1}{\alpha_1^3} - \frac{1}{\alpha_1^1}\right) + \frac{\alpha_2^2}{2\alpha_1^4} - \frac{\alpha_3^1}{3\alpha_1^3}, \tag{7.16}$$

these approximations are very good when $Q_2^* \geq Cond(D_2^d)$, and for other situations we suggest to use the approximation algorithm at the end of section 2.2.

ad (4) Given (A_1, D_1), (A_2^r, D_2^r) and (T_2, O_2) we now determine the superposed demand process, (A_1^d, D_1^d). To find the first two moments of A_1^d and D_1^d we apply the stationary interval method, developed by Whitt (1982), to superpose renewal processes. It is well known that the superposition of two independent renewal processes is itself a renewal process if and only if both processes are Poisson (see Cinlar (1975)). Since the process (T_2, O_2) is not Poisson this situation does not apply here, therefore we are interested in approximations.

The key idea in the stationary interval method is to assume that all renewal processes, including the superposed process, are stationary at time zero. Then the first renewal time of the superposed process is distributed as the first order statistic of the first renewal time of the individual renewal processes. Since these individual renewal processes are stationary at time zero, the first renewal times are distributed as the asymptotic residual life time distribution. The first two moments of A_1^d are found by recursively superposing the currently obtained superposed renewal process with a renewal process not yet included.

We summarize the stationary-interval method for n independent renewal processes below. Let $\{T_i, O_i\}_{i=1}^n$ denote the individual compound renewal processes. Denote (T_0, O_0) the superposed compound renewal process. Then the c.d.f. and mean of T_0 satisfy (see Whitt (1982))

$$\mathbb{E}T_0 = \frac{1}{\sum_{i=1}^{n} \frac{1}{\mathbb{E}T_i}}, \tag{7.17}$$

and

$$1 - F_{T_0}(x) = \sum_{i=1}^{n} \frac{\mathbb{E}T_0}{\mathbb{E}T_i}(1 - F_{T_i}(x)) \prod_{j,j\neq i} \frac{1}{\mathbb{E}T_i} \int_x^\infty (1 - F_{T_j}(y))dy. \tag{7.18}$$

Then for $k \geq 2$ (see Whitt(1979), Appendix 5)

$$\mathbb{E}T_0^k = k(k-1)\mathbb{E}T_0(\prod_{i=1}^{n} \frac{1}{\mathbb{E}T_i}) \int_0^\infty x^{k-2} \prod_{i=1}^{n} \int_x^\infty (1 - F_{T_i}(y))dydx. \tag{7.19}$$

The second moment of T_0 can be computed recursively by superposing two processes at a time. For the situation that the distribution functions of T_i are ME-distributed we derive closed form expressions in Appendix 7.A.

The first two moments of O_0 can be obtained straightforwardly by taking the weighted sum of the individual order sizes.

$$\mathbb{E}O_0 = \sum_{i=1}^{n} \frac{\mathbb{E}T_0}{\mathbb{E}T_i}\mathbb{E}O_i, \tag{7.20}$$

$$\mathbb{E}O_0^2 = \sum_{i=1}^{n} \frac{\mathbb{E}T_0}{\mathbb{E}T_i}\mathbb{E}O_i^2. \tag{7.21}$$

ad (5) Now all ingredients for calculating the optimal reorder point for s_1^* stockpoint 1 are available. Note that $P_{2,1}$ is independent of s_2, since s_2 does not influence (T_2, O_2). Using (A_1^d, D_1^d), L_1, Q_1^* and $P_{2,1,target}$ determine s_1^* by solving s_1 from

$$\frac{\mathbb{E}(D(0, L_1) + U_1 - s_1)^+ - \mathbb{E}(D(0, L_1) + U_1 - s_1 - Q_1^*)^+}{Q_1^*} = 1 - P_{2,1,target}.$$

ad (6) To derive the actual lead times from stockpoint 1 to stockpoint 2 the waiting time characteristics at stockpoint 1 are required. First consider constant transportation times l. Now we must distinguish between the situations $s \geq 0$ and $s < 0$.

Theorem 7.1 *Consider an (s,Q) inventory system with constant lead times l.*

$$\mathbb{P}(W \leq y) = 1 - \frac{\mathbb{E}(D(0,l-y)+D-s)^{+} - \mathbb{E}(D(0,l-y)+D-s-Q)^{+}}{Q} \quad 0 \leq y < l,$$

$$\mathbb{P}(W = l) = \frac{\mathbb{E}(D-(s)^{+})^{+} - \mathbb{E}(D-(s)^{+}-Q)^{+}}{Q},$$

and

$$\mathbb{P}(l < W \leq y) = \frac{\mathbb{E}D(y-l) - \mathbb{E}(D(y-l)+s)^{+}}{Q} \quad y > l, s < 0.$$

For the proof see Appendix 7.B.

By using Theorem 7.1 the first two moments of W can be determined. First consider $s \geq 0$. Using Theorem 7.1 yields

$$\begin{aligned}
\mathbb{E}W &= \int_0^\infty \mathbb{P}(W > y)dy \\
&= \int_0^l \frac{\mathbb{E}(D(0,l-y)+D-s-Q)^{+} - \mathbb{E}(D(0,l-y)+D-s)^{+}}{Q}dy \\
&= l\int_0^l \frac{\mathbb{E}(D(0,y)+D-s-Q)^{+} - \mathbb{E}(D(0,y)+D-s)^{+}}{Q}1/l\,dy \\
&= l\frac{\mathbb{E}(D(0,\tilde{L})+D-s-Q)^{+} - \mathbb{E}(D(0,\tilde{L})+D-s)^{+}}{Q},
\end{aligned} \tag{7.22}$$

where $\tilde{L}$ is uniformly distributed over $(0,l)$. Analogously, we find

$$\mathbb{E}W^2 = l^2\frac{\mathbb{E}(D(0,\hat{L})+D-s-Q)^{+} - \mathbb{E}(D(0,\hat{L})+D-s)^{+}}{Q}, \tag{7.23}$$

where

$$F_{\hat{L}}(t) = \begin{cases} 0 & t < 0; \\ \int_0^t \frac{2(l-y)}{l^2}dy & t < l; \\ 1 & t > l. \end{cases}$$

Note that both $\tilde{L}$ as $\hat{L}$ have substantial probability mass near zero. Therefore we have to be careful with applying (2.14), (2.15) with the asymptotic expressions (2.20) and (2.21) for computing the first two moments of $D(0,\tilde{L})$ and $D(0,\hat{L})$ which are needed to compute both moments of W.

For stochastic L we have to condition on $L = l$ and apply Theorem 7.1. We find for the first two moments of W (see Appendix 7.C, or de Kok (1993)).

$$\mathbb{E}W = \mathbb{E}L\frac{\mathbb{E}(D(0,\tilde{L})+D-s-Q)^+ - \mathbb{E}(D(0,\tilde{L})+D-s)^+}{Q}, \tag{7.24}$$

$$\mathbb{E}W^2 = \mathbb{E}L^2\frac{\mathbb{E}(D(0,\hat{L})+D-s-Q)^+ - \mathbb{E}(D(0,\hat{L})+D-s)^+}{Q}, \tag{7.25}$$

where

$$F_{\tilde{L}}(x) = \frac{1}{\mathbb{E}L}\int_0^x (1-F_L(y))dy,$$

$$F_{\hat{L}}(x) = \frac{2}{\mathbb{E}L^2}\int_0^x\int_y^\infty (z-y)dF_L(z)dy.$$

Note that if L is deterministic the two expressions for $F_{\tilde{L}}(x)$ and $F_{\hat{L}}(x)$ coincide. For $s < 0$ analoguous results can be derived.

ad (7) Now we can compute the optimal reorder point for stockpoint 2. Using (A_2^d, D_2^d), L_2, W_1, Q_2^* and $P_{2,2,target}$ determine s_2^* by solving s_2 from

$$\frac{\mathbb{E}(D(0,L_2+W_1)+U_2-s_2)^+ - \mathbb{E}(D(0,L_2+W_1)+U_2-s_2-Q_2^*)^+}{Q_2^*} = 1 - P_{2,2,target}.$$

7.1.2 Validation of the procedure

We used discrete event simulation to validate the quality of the approximations in terms of the deviation of the calculated performance measures by the algorithm, described in the previous section, and the performance measures computed by simulation. By using formulas (7.2) for the P_2 service level we implicitly assume that replenishment orders arrive in one batch (also in case of backlogging) and that customer orders are splitted when not enough stock is available to fulfil the complete customer order. Yet, we measure at stockpoint 1 as if all the deliveries are splitted. In the simulation experiments, replenishment orders of stockpoint 2 arriving at stockpoint 1 are not splitted when shortages occur, however, the available physical stock is reserved for this specific order. On the other hand, external customers are splitted. We investigated the performance in three experiments (e_1, e_2, e_3) where we varied D_{max} between 5, 10, 20, 40, and ∞. The input values for each of the three experiments are given in Table 7.1. For each of the experiments we calculated s_1^*, Q_1^*, s_2^* and Q_2^* by using the heuristic above. We simulated the system for 500.000 time units. Furthermore, the demand sizes, interarrival times, and the lead times, are ME distributed. We computed $B_{4,i}$, $\mathbb{E}A_1^d$ and $\mathbb{E}T_i$ with the heuristic and compared them with the simulation output. Table 7.2 shows the absolute and relative errors for each experiment. The

Table 7.1: Basic parameter setting for the validation

	i	$(\mathbb{E}A_i, c_{A_i})$	$(\mathbb{E}D_i, c_{D_i})$	$(\mathbb{E}L_i, c_{L_i})$	$P_{2,i,target}$	a_i	b_4
e_1	1	(10,1)	(150,0.75)	(10,0)	0.90	500	0.02
	2	(1,1)	(5,0.75)	(5,0)	0.90	100	0.02
e_2	1	(5,1)	(50,1)	(20,0.5)	0.95	500	0.02
	2	(1,1)	(5,1)	(5,0.5)	0.95	100	0.02
e_3	1	(1,1)	(25,1)	(10,0.75)	0.99	500	0.02
	2	(1,1)	(5,2)	(5,0.75)	0.99	100	0.02

Table 7.2: The mean absolute and mean relative errors

	$P_{2,i}$		$\mathbb{E}A_0$		$B_{4,i}$		$\mathbb{E}T_i$	
	abs.	rel.	abs.	rel.	abs.	rel.	abs.	rel.
e_1	0.0077	7,7 %	0.0176	0.20 %	3.98	0.99 %	0.192	0.36 %
e_2	0.0068	13.6 %	0.0104	0.25 %	1.94	0.38 %	0.1107	0.16 %
e_3	0.00056	5.6 %	0.0023	0.25 %	0.79	0.11 %	0.0995	0.15 %

comparison of $P_{2,i}$, $B_{4,i}$ and $\mathbb{E}T_i$ is based on averaging the outcomes of the comparison for the 5 different values of D_{max} as well averaging over the two stockpoints. From these results follows that derivations are small in most cases.

7.1.3 Numerical results

In this section we investigate the effects of large order overflow on the total relevant costs. With the heuristic algorithm of section 7.1.1 we compute control parameters, the performance measures and the relevant costs. To give an impression of the ordering, holding and transportation costs we plotted these costs separately in Figure 7.2 as function of D_{max}, although summing over both stockpoints. The input values for this experiment are given in Table 7.3. From Figure 7.2 we see that both the holding costs and the ordering costs

Table 7.3: The basic parameters setting for the numerical examples

i	$(\mathbb{E}A_i, c_{A_i})$	$(\mathbb{E}D_i, c_{D_i})$	$(\mathbb{E}L_i, c_{L_i})$	$P_{2,i,target}$	a_i	b_4	k_i
1	(5,1)	(100,1)	(20,0.5)	0.90	500	0.05	40
2	(1,1)	(10,2)	(5,0.5)	0.90	100	0.05	20

decrease when large order overflow is applied; for smaller values of D_{max} the savings increase. The transportation costs, on the other hand, increase when large order overflow is

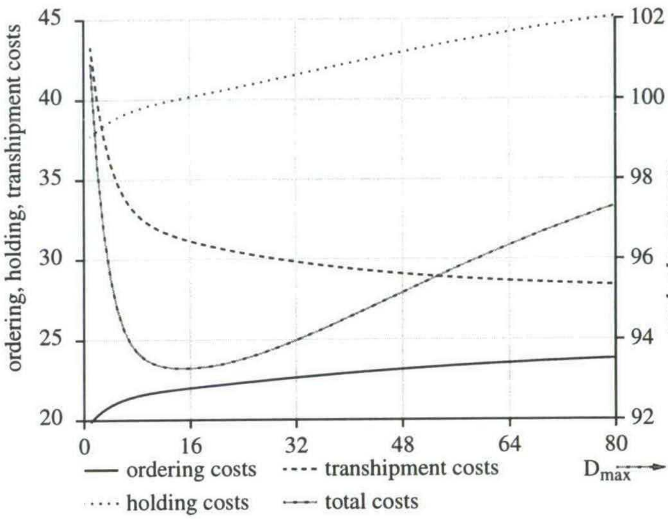


Figure 7.2: The relevant cost as function of D_{max}

applied. For $D_{max} \to \infty$ the transportation cost are equal to $\sum_{i=1}^{2} \frac{k_i}{EA_i}$ (which equals 28 in the example) and for $D_{max} \to 0$ the transportation cost are equal to $k_1 \sum_{i=1}^{2} 1/\mathbb{E}A_i$ (which equals 48 in the example). Moreover, we see that the total cost are indeed convex in this example and it has its minimal value at $D_{max} = 15$. In our numerical experiments no counter examples were found of the total costs being convex.

Now we focus on the effect of the transportation costs k_1 and k_2. Since $\sum_{i=1}^{2} \frac{k_i}{EA_i}$ is invariant under large order overflow, we only have to consider the differences between k_1 and k_2 (see 7.1). In case $(k_1 - k_2) \leq 0$ using stockpoint 2 is only useful if transportations from stockpoint 1 to customers in the neighbourhood of stockpoint 2 is not possible for some reason (e.g. long transportation times, taxes). In this evaluation we assume that customers do not suffer any disadvantages when orders are re-routed. Under these circumstances only situations are worthwhile considering where $(k_1 - k_2) > 0$, hence using stockpoint 2 is based only on cost considerations, e.g. co-ordinated transportation. The input values for this experiment are again given in Table 7.3. However, k_1 is fixed to 100 and k_2 is varied between $0, 20, \ldots, 100$. Figure 7.3 shows the total relevant costs $C(s_1^*, Q_1^*, s_2^*, Q_2^*, D_{max})$ as function of D_{max}, where s_1^*, Q_1^*, s_2^* and Q_2^* are determined with the algorithm proposed in section 7.1.1 . We see that the optimal value of D_{max} decreases when $k_1 - k_2$ decreases. Moreover for $k_1 - k_2 = 0$ ($k_2 = 100$) we see that stockpoint 2 indeed becomes obsolete, i.e. D^*_{max} equals zero. This means that all customers arriving at stockpoint 2 are re-routed to stockpoint 1. Only for the case that $k_1 - k_2 = 100$ ($k_2 = 0$) large order is not profitable. Obviously, the price to pay for diverting a customer, $k_2 - k_1$, is too high. Hence, from the large order overflow model we can not only conclude whether or not large order overflow is profitable, but also we can draw conclusions whether a second stockpoint is profitable (but only when there are no lead time restrictions).

In the next example, we consider the optimal value D_{max}. Here we investigate the influence of the target service level $P_{2,i,target}$ and the carrying cost per unit per unit of time (b_4). The input values for this example are again given in Table 7.3. Furthermore, $k_1 - k_2 = 50$, $P_{2,1,target} = P_{2,i,target}$ and are varied between 0.80 and 0.999, and b_4 is varied between 0.02, 0.03, 0.04 and 0.05. Figure 7.4 shows the the optimal value for D_{max}. From these curves it follows that D^*_{max} decreases when the service level increases or when b_4 increases. This means that for high service levels or high carrying cost per unit per unit of time large order overflow is more likely to be profitable.

In the final example of this section, we investigate the effects of k_2 and a_2 on D^*_{max} and Q^*_2. The input values for this example are again given in Table 7.3. Furthermore, a_2 is varied between 10, 50, 100, 200, 300, 400, and 500, and $k_1 = 40$. Figure 7.5 and 7.6 show D^*_{max} and Q^*_2 respectively as function of k_2, where $k_2 \in [0, k_1]$. The optimal value for D_{max} decreases when k_2 increases confirming an the earlier example. Moreover, for k_2 larger than a certain threshold value k^0_2 stockpoint 2 becomes obsolete. The value for k^0_2 decreases when a_2 increases, which means that stockpoint 2 becomes redundant for large values of a_2 even when $(k_1 - k_2)$ is large. The optimal value for Q_2 decreases when when k_2 increases. The reason for this is that D^*_{max} declines and therefore the total demand delivered by the stockpoint 2 decreases.

Moreover, it is possible that Q^*_2 is non-increasing with the value of a_2. At first sight this might be counter-intuitive The reason for this is that D^*_{max} for large values of a_2 is lower than for low values of a_2 (i.e. the total demand delivered by stockpoint 2 is lower).

7.2 The two-echelon system with N stockpoints

Consider a divergent multi-echelon system with one central depot supplying N local stockpoints. The central depot is denoted as stockpoint 0. The local stockpoints i ($i = 1, \ldots, N$) face independent compound Poisson demand processes with interarrival times A_i and demand sizes D_i. Only at the local stockpoints large order overflow is applied. Customers arriving at local stockpoint i with demand larger than $D_{max,i}$ are re-routed to stockpoint 1 (see Figure 7.7). The replenishments at stockpoint i ($i = 0, \ldots, N$) are controlled by an (s_i, Q_i) policy. The inter-replenishment times and replenishment sizes initiated at stockpoint i are denoted by (T_i, O_i). The deliveries from stockpoint i ($i = 0, \ldots, N$) are described by the demand process (A^d_i, D^d_i) whereas the re-routed customers of stockpoint i are described by the demand process (A^r_i, D^r_i). The transportation time to stockpoint i is a random variable L_i. Demand which cannot be delivered directly from shelf is backordered. The problem now is to find the control variables s_i, Q_i, $D_{max,i}$ for each stockpoint i. The objective is to minimize the sum of the ordering, holding and tranportation costs subject

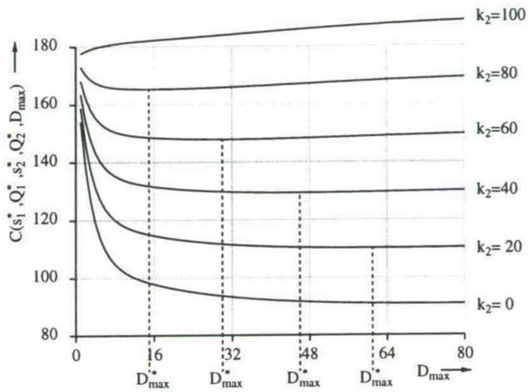


Figure 7.3: The total relevant cost as function of k_2 and D_{max}

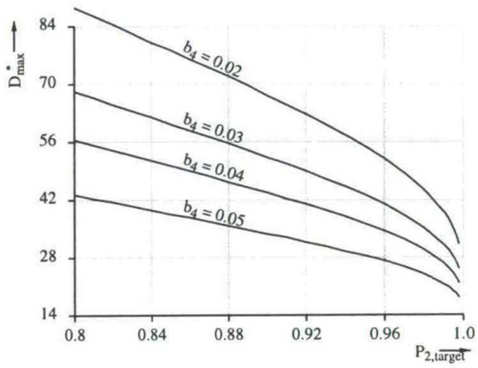


Figure 7.4: The optimal value of D_{max} as function of $P_{2,target}$ and b_4

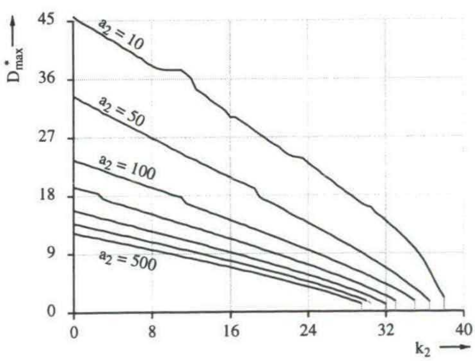


Figure 7.5: The optimal value of D_{max} as function of a_2 and k_2

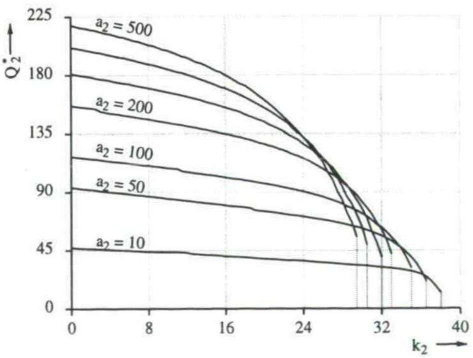


Figure 7.6: The optimal value of Q as function of a_2 and k_2

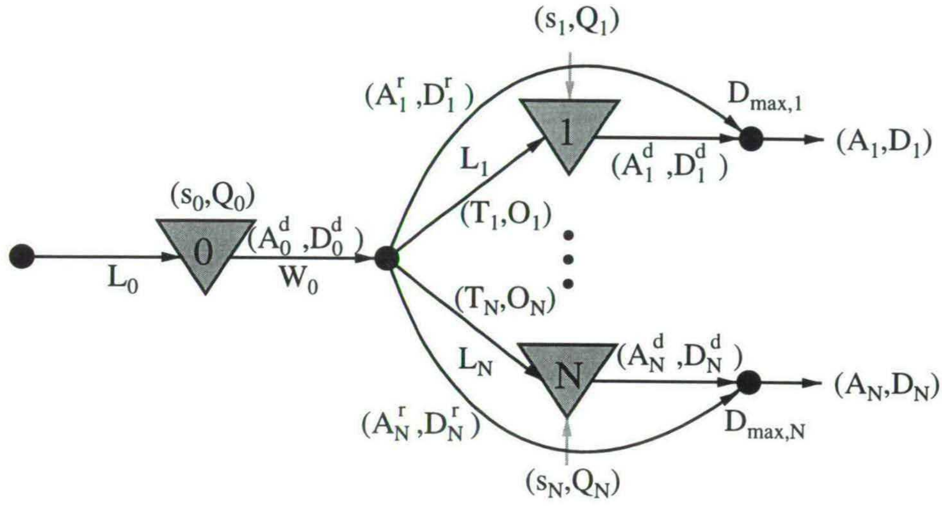


Figure 7.7: The 2-echelon network

to a service level constraint. Each stockpoint i has to guarantee a certain P_2 performance level denoted by $P_{2,i,target}$.

Ordering costs are fixed per replenishment and are denoted by a_i at stockpoint i. The transportation from stockpoint i to external customers are fixed per transportation, i.e. independent of the size, and are denoted by k_i. We assume that the transportation costs of replenishments to stockpoint i are included in a_i. The inventory holding costs are proportional to the average physical stock with rate b_4 at each stockpoint.

Let $B_{4,i}$ be the average physical stock at stockpoint i and $\mathbf{s} = (s_0, \ldots, s_N)$, $\mathbf{Q} = (Q_0, \ldots, Q_N)$ and $\mathbf{D_{max}} = (D_{max,1}, \ldots, D_{max,N})$, then the total relevant costs as function of the control parameters is given by the following expression

$$C(\mathbf{s}, \mathbf{Q}, \mathbf{D_{max}}) = \sum_{i=0}^{N} \Big(\frac{a_i}{\mathbb{E}T_i} + b_4 B_{4_i} + \frac{k_i}{\mathbb{E}A_i^d} + \frac{k_1}{\mathbb{E}A_i^r} \Big) \tag{7.26}$$

Now we formulate the following minimization problem

$$\begin{array}{ll} (\mathcal{P}_2) & \text{minimize} \quad C(\mathbf{s}, \mathbf{Q}, \mathbf{D_{max}}) \\ & \text{s.t.} \quad P_{2,i} = P_{2,i,target} \quad (i = 0, \ldots, N). \end{array}$$

For solving this optimization problem we can apply a similar heuristic approach as presented in section 7.1.1.

(1) Based on (A_i, D_i) and $D_{max,i}$ determine (A_i^d, D_i^d) and (A_i^r, D_i^r) for $(i = 1, \ldots, N)$.

(2) Determine Q_i^* for $i = 0, \ldots, N$ by applying the economic order quantity.

(3) Use (A_i^d, D_i^d) and Q_i^* to determine (T_i, O_i) for $(i = 1, \ldots, N)$.

(4) Use (A_0, D_0), (A_i^r, D_i^r) and (T_i, O_i) for $(i = 1, \ldots, N)$ to determine (A_1^d, D_1^d).

(5) Using (A_0^d, D_0^d), L_0, Q_0^* and $P_{2,0,target}$ determine s_0^* by solving s_0 from the service level constraint.

(6) Using (A_0^d, D_0^d), L_0, Q_0^* and s_0 determine W_0.

(7) For $(i = 2, \ldots, N)$ use (A_i^d, D_i^d), L_i, W_0, Q_i^* and $P_{2,i,target}$ to determine s_i^* by solving s_i from the service level constraint.

7.2.1 Numerical results

In this section we investigate the effects of large order overflow on the total relevant costs in the 2-echelon system. With the heuristic algorithm presented above we computed for given values of $D_{max,i}$ the optimal control parameters s_i^* and Q_i^* for $(i = 1, \ldots, N)$, the performance measures and the relevant costs. The input values for the example are given in Table 7.4. The optimal control parameters s_i and Q_i are computed for $D_{max,1} = \infty$ and

Table 7.4: Parameter setting for the 2-echelon example

	$(\mathbb{E}A_i, c_{A_i})$	$(\mathbb{E}D_i, c_{D_i})$	$(\mathbb{E}L_i, c_{L_i})$	$P_{2,i,target}$	a_i	b_4	k_i
0	(0.00, 0.00)	(0.00, 0.00)	(10.00, 0.5)	0.95	500	0.05	20
1	(10.00, 0.20)	(1.00, 1.00)	(5.00, 1.0)	0.95	100	0.05	10
2	(10.00, 0.60)	(1.00, 1.00)	(5.00, 0.9)	0.95	100	0.05	10
3	(10.00, 1.00)	(1.00, 1.00)	(5.00, 0.8)	0.95	100	0.05	10
4	(10.00, 1.40)	(1.00, 1.00)	(5.00, 0.7)	0.95	100	0.05	10
5	(10.00, 1.80)	(1.00, 1.00)	(5.00, 0.6)	0.95	100	0.05	10
6	(10.00, 0.40)	(1.00, 1.00)	(10.00, 0.1)	0.95	100	0.05	10
7	(10.00, 0.80)	(1.00, 1.00)	(10.00, 0.2)	0.95	100	0.05	10
8	(10.00, 1.20)	(1.00, 1.00)	(10.00, 0.3)	0.95	100	0.05	10
9	(10.00, 1.60)	(1.00, 1.00)	(10.00, 0.4)	0.95	100	0.05	10
10	(10.00, 2.00)	(1.00, 1.00)	(10.00, 0.5)	0.95	100	0.05	10

two values of $D_{max,i}$ for $i = (2, \ldots, 11)$, namely 10 and ∞. In Table 7.5 we plotted for each depot i the optimal control parameters (s_i^* and Q_i^*), the associated ordering costs $(C_{o,i})$, the holding costs $(C_{h,i})$, the transportation costs $(C_{t,i})$ and the total cost (C_i).

Table 7.5: Optimal control and the associated costs for the 2-echelon example

i	$D_{max,i}$	s_i^*	Q_i^*	$C_{o,i}$	$C_{h,i}$	$C_{t,i}$	C_i
0	∞	1531.9	1414.2	35.36	62.92	68.95	167.22
1	10.00	38.4	133.7	3.34	4.16	5.27	12.77
2	10.00	25.8	117.4	2.93	3.37	5.77	12.08
3	10.00	17.8	102.8	2.57	2.80	6.32	11.69
4	10.00	13.8	95.1	2.38	2.50	7.03	11.90
5	10.00	12.0	92.4	2.31	2.37	7.67	12.35
6	10.00	46.6	125.6	3.14	3.49	5.52	12.15
7	10.00	34.9	108.2	2.71	2.98	5.99	11.67
8	10.00	29.1	98.1	2.45	2.70	6.67	11.82
9	10.00	26.9	93.4	2.33	2.59	7.36	12.29
10	10.00	27.2	92.0	2.30	2.60	7.93	12.83
0	∞	1730.9	1414.2	35.36	72.99	0.00	108.35
1	∞	97.8	200.0	5.00	7.41	10.00	22.41
2	∞	95.5	200.0	5.00	7.27	10.00	22.27
3	∞	98.2	200.0	5.00	7.40	10.00	22.40
4	∞	106.9	200.0	5.00	7.83	10.00	22.83
5	∞	122.5	200.0	5.00	8.60	10.00	23.60
6	∞	121.3	200.0	5.00	6.00	10.00	21.00
7	∞	135.8	200.0	5.00	6.74	10.00	21.74
8	∞	159.0	200.0	5.00	7.92	10.00	22.92
9	∞	189.7	200.0	5.00	9.48	10.00	24.48
10	∞	227.0	200.0	5.00	11.36	10.00	26.35

For $D_{max,i} = 10$ we find for the total system costs $(\sum_{i=0}^{10} C_i)$: 288.78, while for $D_{max,i} = \infty$ this amounts to 338.33. Hence, for this example large order overflow is profitable. For the local stockpoints $i = 1, \ldots, 11$, s_i^* and Q_i^* decrease significantly, and even s_1^* decreases due to the decreases of Q_i^*. Yet Q_1^* remain constant because the total demand delivered by stockpoint 0 remains constant.

7.3 Conclusions and future research

In this chapter we considered the profitability of large order overflow in a divergent two-echelon network, where stockpoints are controlled by (s, Q) policies. A heuristic algorithm is proposed for deriving the optimal control parameters s_i and Q_i for all the stockpoints, given values of r_i and $D_{max,i}$. In the model discussed in this chapter we modelled demand processes as compound Poison processes. The extension to compound renewal process is clearly a topic for future research.

A trade-off between a decrease in ordering and holding costs versus an increase in transportation costs, enables management to decide to use large order overflow or not. Moreover, it is possible for management to investigate the costs of using a certain stockpoint i by evaluating the situation $D_{max,i} = 0$. Note that lead time restrictions for external customers may prevent this.

From the numerical examples it turns out that large order overflow is profitable in many cases. Only when the differences in transportation between the original stockpoint and the alternative stockpoint become too large, large order overflow is not profitable.

The two-echelon network with N local stockpoints, $(\mathcal{P}_2)$, clearly is a $3N$ dimensional optimization problem. The heuristic presented in section 7.2 solves $(\mathcal{P}_2)$ for given values for $D_{max,i}$ for $i = 1, \ldots, N$. For finding the optimal control parameters $D^*_{max,i}$ we can, for example, use the downhill simplex method in multidimensions.

For a general multi-echelon network, r_i is also a control variable which must be optimized. For a network with N stockpoints that are $N(N-1)$ possible combinations for $r_1, \ldots, r_N$, leaving for each combination a $3N$ dimenisonal optimization problem. Not every combination for $r_1, \ldots, r_N$ is valid in a practical setting. For example, cycles should be avoided, but lateral transhipments could be considered. In this case we could use simulated annealing or taboo search.

Appendix 7.A: The second moment of the superposed renewal process

Only two processes are superposed at a time. We consider here the situation that T_1 and T_2 are both ME-distributed, to be more precise

$$\begin{aligned}
1 - F_{T_1}(x) &= \sum_{j=1}^{2} p_j \sum_{t=0}^{k_j-1} \frac{(\mu_j x)^t}{t!} e^{-\mu_j x}, \\
1 - F_{T_2}(x) &= \sum_{i=1}^{2} q_i \sum_{s=0}^{l_i-1} \frac{(\rho_i x)^s}{s!} e^{-\rho_i x}.
\end{aligned}$$

Using (7.19) with $k = 2$ yields

$$\begin{aligned}
\mathbb{E}T_0^2 &= \frac{2\mathbb{E}T_0}{\mathbb{E}T_1\mathbb{E}T_2} \int_0^\infty \prod_{i=1}^{2} \int_x^\infty (1 - F_{T_i}(y))dydx \\
&= \frac{2\mathbb{E}T_0}{\mathbb{E}T_1\mathbb{E}T_2} \int_0^\infty \Big(\int_x^\infty \sum_{j=1}^{2} p_j \sum_{t=0}^{k_j-1} \frac{(\mu_j y)^t}{t!} e^{-\mu_j y} dy \Big) \Big(\int_x^\infty \sum_{i=1}^{2} q_i \sum_{s=0}^{l_i-1} \frac{(\rho_i y)^s}{s!} e^{-\rho_i y} dy \Big) dx \\
&= \frac{2\mathbb{E}T_0}{\mathbb{E}T_1\mathbb{E}T_2} \int_0^\infty \Big(\sum_{j=1}^{2} p_j \sum_{t=0}^{k_j-1} \frac{1}{\mu_j} \int_x^\infty \mu_j^{(t+1)} \frac{y^t}{t!} e^{-\mu_j y} dy \Big) \Big(\sum_{i=1}^{2} q_i \sum_{s=0}^{l_i-1} \frac{1}{\rho_i} \sum_{n=0}^{s} \frac{(\rho_i x)^n}{n!} e^{-\rho_i x} \Big) dx \\
&= \frac{2\mathbb{E}T_0}{\mathbb{E}T_1\mathbb{E}T_2} \int_0^\infty \Big(\sum_{j=1}^{2} p_j \sum_{t=0}^{k_j-1} \frac{(k_j - t)}{\mu_j} \frac{(\mu_j x)^t}{t!} e^{-\mu_j x} \Big) \Big(\sum_{i=1}^{2} q_i \sum_{s=0}^{l_i-1} \frac{(l_i - s)}{\rho_i} \frac{(\rho_i x)^s}{s!} e^{-\rho_i x} \Big) dx \\
&= \frac{2\mathbb{E}T_0}{\mathbb{E}T_1\mathbb{E}T_2} \sum_{j=1}^{2} \sum_{i=1}^{2} p_j q_i \sum_{t=0}^{k_j-1} \sum_{s=0}^{l_i-1} \frac{(k_j - t)}{\mu_j} \frac{(l_i - s)}{\rho_i} \int_0^\infty \frac{(\mu_j x)^t}{t!} \frac{(\rho_i x)^s}{s!} e^{-(\mu_j + \rho_i)x} dx \\
&= \frac{2\mathbb{E}T_0}{\mathbb{E}T_1\mathbb{E}T_2} \sum_{j=1}^{2} \sum_{i=1}^{2} p_j q_i \sum_{t=0}^{k_j-1} \sum_{s=0}^{l_i-1} (k_j - t)(l_i - s) \binom{t+s}{t} \frac{\mu_j^{t-1} \rho_i^{s-1}}{(\mu_j + \rho_i)^{t+s+1}}
\end{aligned}$$

Appendix 7.B: The c.d.f. of the customer waiting times for constant lead times

First consider the case $s \geq 0$. Let $X(t)$ denote the inventory position at time t, and $W(t)$ the waiting time of a customer who arrives at time epoch t and has demand D. This customer has to wait less than y ($y < l$) if and only if $X(t+y-l)$ (which is available to be issued at time $t+y$) minus $D(t+y-l,t)$ is larger than D which implies $X(t) \geq D$. Hence,

$$\mathbb{P}(W(t) \leq y) = \mathbb{P}(X(t+y-l) - D(t+y-l,t) \geq D) \quad y < l.$$

Conditioning on $X(t+y-l)$ and on D, and using that the $X(t+y-l)$ in an (s,Q) system is uniformly distributed over $(s, s+Q)$ yields

$$\begin{aligned}\mathbb{P}(W \leq y) &= 1/Q \int_s^{s+Q} \mathbb{P}(D(0,l-y) + D \leq u)du \\ &= 1 - \frac{\mathbb{E}(D(0,l-y)+D-s)^+ - \mathbb{E}(D(0,l-y)+D-s-Q)^+}{Q}.\end{aligned}$$

A customer waits precisely l periods if and only if $X(t) < D$ Hence,

$$\mathbb{P}(W(t) = l | X(t) = u) = \mathbb{P}(D \geq u | X(t) = u)$$

Conditioning on $X(t)$ and on D and using that the $X(t)$ in an (s,Q) system is uniformly distributed over $(s, s+Q)$ yields

$$\begin{aligned}\mathbb{P}(W = l) &= 1/Q \int_s^{s+Q} \int_u^{\infty} dF_D(x)du \\ &= \frac{\mathbb{E}(D-s)^+ - \mathbb{E}(D-s-Q)^+)}{Q}\end{aligned}$$

Now consider $s < 0$. Again consider a customer who arrives at time epoch t and has demand D. As for the case $s \geq 0$, this customer has to wait less than y ($y < l$) if and only if $X(t+y-l)$ minus $D(t+y-l,t)$ is larger than D which again implies $X(t) \geq D$. Then,

$$\mathbb{P}(W(t) \leq y) = \mathbb{P}(D(t+y-l,t) + D \leq X(t+y-l)) \quad y < l.$$

Conditioning on $X(t+y-l)$ where $X(t+y-l) > 0$ and on D and using that the $X(t+y-l)$ in an (s,Q) system is uniformly distributed over $(s, s+Q)$ yields

$$\begin{aligned}\mathbb{P}(W \leq y) &= 1/Q \int_0^{s+Q} \mathbb{P}(D(l-y) + D \leq u)du \\ &= 1/Q\mathbb{E}(s+Q-D(l-y)-D)^+ \\ &= 1 - \frac{\mathbb{E}(D(l-y) + \mathbb{E}D - s - \mathbb{E}(D(l-y)+D-(s+Q))^+}{Q}.\end{aligned}$$

A customer waits precisely l periods if and only if $X(t) - s$ is less than D Hence,

$$\mathbb{P}(W(t) = l | X(t) = u) = \mathbb{P}(D \geq u - s | X(t) = u)$$

Conditioning on $X(t)$ and on D and using that the $X(t)$ in an (s, Q) system is uniformly distributed over $(s, s+Q)$ yields

$$\begin{aligned}
\mathbb{P}(W = l) &= 1/Q \int_s^{s+Q} (1 - F_D(u - s))du \\
&= 1/Q \int_0^Q (1 - F_D(u))du \\
&= \frac{\mathbb{E}D - \mathbb{E}(D - Q)^+}{Q}
\end{aligned}$$

For $s < 0$, we have also a possibility that $W(t) > l$. This occurs when a customer arrives at time t and finds $s < X(t) - D < 0$, where $X(t) - D$ is the inventory position just after the customer arrival. Such a customer would have to wait until the next order placement plus l. Hence,

$$\mathbb{P}(l < W_t \leq y | X(t) - D = u) = \mathbb{P}(D(y - l) \geq u - s)$$

Conditioning on $X(t) - D$ and using that the $X(t) - D$ in an (s, Q) system is uniformly distributed over $(s, s+Q)$ yields

$$\begin{aligned}
\mathbb{P}(l < W \leq y) &= 1/Q \int_s^0 (1 - F_{D(y-l)}(u - s))du \\
&= 1/Q \int_0^{-s} (1 - F_{D(y-l)}(u))du \\
&= \frac{\mathbb{E}D(y - l) - \mathbb{E}(D(y - l) - (-s))^+}{Q}
\end{aligned}$$

□

Appendix 7.C: The moments of the customer waiting times for stochastic lead times

By conditioning on $L = l$ we find

$$\begin{aligned}
\mathbb{E}W &= \int_0^\infty \mathbb{P}(W > y)dy \\
&= \int_0^\infty \int_0^l \mathbb{P}(W > y | L = l) dy dF_L(l) \\
&= \int_0^\infty \int_0^l \frac{\mathbb{E}(D(0, l-y) + D - s - Q)^+ - \mathbb{E}(D(0, l-y) + D - s)^+}{Q} dy dF_L(l) \\
&= \int_0^\infty \int_0^l \frac{\mathbb{E}(D(0, z) + D - s - Q)^+ - \mathbb{E}(D(0, z) + D - s)^+}{Q} dz dF_L(l) \\
&= \int_0^\infty \int_z^\infty \frac{\mathbb{E}(D(0, z) + D - s - Q)^+ - \mathbb{E}(D(0, z) + D - s)^+}{Q} dF_L(l) dz \\
&= \int_0^\infty \frac{\mathbb{E}(D(0, z) + D - s - Q)^+ - \mathbb{E}(D(0, z) + D - s)^+}{Q} (1 - F_L(z)) dz \\
&= \mathbb{E}L \frac{\mathbb{E}(D(0, \tilde{L}) + D - s - Q)^+ - \mathbb{E}(D(0, \tilde{L}) + D - s)^+}{Q},
\end{aligned}$$

where

$$f_{\tilde{L}}(z) = 1/\mathbb{E}L(1 - F_L(z)) \quad x \geq 0.$$

For the second moment an analoguous approach can be followed, namely

$$\begin{aligned}
\mathbb{E}W^2 &= \int_0^\infty 2y\mathbb{P}(W > y)dy \\
&= \int_0^\infty \int_0^l 2y\mathbb{P}(W > y | L = l) dy dF_L(l) \\
&= \int_0^\infty \int_0^l 2y \frac{\mathbb{E}(D(0, l-y) + D - s - Q)^+ - \mathbb{E}(D(0, l-y) + D - s)^+}{Q} dy dF_L(l) \\
&= \int_0^\infty \int_0^l 2(l-z) \frac{\mathbb{E}(D(0, z) + D - s - Q)^+ - \mathbb{E}(D(0, z) + D - s)^+}{Q} dz dF_L(l) \\
&= \int_0^\infty \int_z^\infty 2(l-z) \frac{\mathbb{E}(D(0, z) + D - s - Q)^+ - \mathbb{E}(D(0, z) + D - s)^+}{Q} f_L(l) dl dz
\end{aligned}$$

$$= \quad \mathbb{E}L^2 \frac{\mathbb{E}(D(0,\hat{L}) + D - s - Q)^+ - \mathbb{E}(D(0,\hat{L}) + D - s)^+}{Q},$$

where

$$f_{\hat{L}}(x) = 1/\mathbb{E}L^2 \int_z^\infty 2(l - z) dF_L(l) \quad z \geq 0.$$

□

Chapter 8

Delivery splitting

This chapter is based on Janssen, de Kok, and van der Duyn Schouten (1995). The issue of splitting a single order in a number of subsequent (equally sized) deliveries in the same reorder cycle has received much attention during the last couple of years (cf. section 1.4 and 1.4.2). Most of the research in this area is focused on the analysis of splitting an order over a number of different suppliers (see Chapter 6 of this thesis). More recently, also order splitting in a single supplier context has been studied by Chiang and Chiang (1996). All these studies approach the order splitting concept from the perspective of the buyer.

The motivating question for this research is: "To what extent can order splitting contribute to safety stock reduction on the buyers side?" The answer to this question is encouraging. Chiang and Chiang (1996) report that by splitting a single order (generated according to an optimal (s,Q) ordering policy) into two equally sized deliveries can yield cost savings up to 20% under realistic settings of cost parameters, while splitting in three equally sized deliveries can give additional savings of 10%.

In this chapter delivery splitting is analysed from the suppliers perspective. We will provide an analytical (approximation) approach to analyse the delivery splitting concept. We will assume that the underlying replenishment policy of the supplier is an (s, Q) policy, where s denotes the reorder point and Q the reorder quantity. A typical aspect of delivery splitting from the suppliers perspective is that through delivery splitting not only the variability of the demand process is decreased, but also actual information about future deliveries is obtained. When it is decided that a (large) order is splitted into a number of equally spaced smaller deliveries than this information can be used in the reorder policy of the supplier. In this chapter we will consider both the situation in which this additional information is used and the situation where it is not. The difference in performance of both situations can be used to decide whether the internal information systems should be adapted to take the additional information into account.

The organization of the chapter is as follows. In section 8.1 we present a detailed description of the delivery splitting model and present the associated optimization problem. As in many inventory systems, the first two moments of the demand during the lead time and the undershoot are very important. For inventory models with delivery splitting these moments can not be determined with the usually applied formulas (2.12), (2.14) and (2.15). Sections 8.2 and 8.3 presents two methods to calculate these moments. The method of section 8.2 is 'quick and dirty', while the method of section 8.3 is more accurate and sophisticated at the expense of increased complexity and computational requirements. Section 8.4 is concerned with a numerical analysis aimed at validation of the approximation approaches of sections 8.2 and 8.3. In section 8.5 we will present an alternative model where information about future deliveries (due to splitting of previously placed orders) is used explicitly in the sense that order decisions are based on 'free inventory position' instead of the inventory position itself In section 8.6 we investigate the solutions of the optimization problem. Also an indication is given of the effects of using delivery splitting and using information about the future deliveries in controlling the system. Finally, section 8.7 presents some conclusions and recommendations for future research.

8.1 Model description

We assume that the demand process is a compound Poisson process with arrival rate λ. The c.d.f. of the demand size is denoted by F_D and the p.d.f. is given by f_D. Replenishment of stock occurs according to a continuous review (s,Q)-policy. The lead time, L, of the replenishment orders is supposed to be deterministic. This guarantees that replenishment orders do not cross in time. For compound renewal demand processes and stochastic lead times the formulas become extremely cumbersome. Therefore we restrict ourselves to compound Poisson demand and deterministic lead times.

Orders which cannot be delivered directly from stock on hand will be backordered. However, large demands will not be delivered in one single batch, even in case the inventory level is sufficiently large. The customer receives only a limited quantity, D_{max}, at a time. If the demand size is larger than D_{max}, starting at the demand epoch, an amount of D_{max} is delivered in a number of shipments which are T time units apart. Consequently, all quantities delivered are equal to D_{max} except possibly the last. When a customer with demand of size D arrives at epoch t this results in the following delivery scheme: deliver D_{max} on epoch $t + jT$, $(0 \leq j < n)$; and deliver $D - nD_{max}$ on epoch $t + nT$ where $n = \max\{m \in \mathbb{N} | mD_{max} \leq D\}$.

The problem is to find optimal values for the control variables s, Q, D_{max} and T. The objective is to minimize the sum of the ordering, holding and transshipment costs subject to a service level constraint. As a service measure the P_2-service measure is used. The target service is denoted by $P_{2,target}$.

We define the following performance measures (cf. section 2.1.4):

$B_4(s, Q, D_{max}, T)$	the expected average physical stock level;
$\lambda^*(s, Q, D_{max}, T)$	the expected number of deliveries to the customers;
$P_2(s, Q, D_{max}, T)$	the fraction of demand delivered directly from shelf.

The holding cost are proportional to the expected average physical stock level, i.e. to stock one unit of product costs b_4 per day. The ordering costs are proportional to the number of replenishment orders, i.e. to place one replenishment order costs a, and the transshipment costs are proportional the number of customer deliveries, i.e. each customer delivery costs k.

The total relevant costs per unit of time, as function of the control parameters, is given by the following expression

$$C(s,Q,D_{max},T) = \frac{a\lambda \mathbb{E}D}{Q} + b_4 B_4(s,Q,D_{max},T) + k\lambda^*(s,Q,D_{max},T)$$

Then we formulate the following minimization problem

$$\begin{aligned} (\mathcal{P}_1) \qquad & \text{minimize} \quad C(s,Q,D_{max},T) \\ & \text{s.t.} \quad P_2(s,Q,D_{max},T) = P_{2,1,target}; \\ & \qquad\;\; Q \geq 0. \end{aligned}$$

Denote by $D(L)$ the total demand during the lead time L and by U the undershoot under the reorder point s of the customers demand that triggers a replenishment. Due to the delivery splitting strategy, the inventory system is confronted with a transformed demand process. This process is strongly correlated, and therefore also the demand in successive replenishment cycles are correlated. On the other hand the correlation between two replenishment cycles will be negligible for large Q, which is exactly the environment where delivery splitting makes sense. Our approximation is to neglect this correlation, and therefore we may use the standard formulas for the (s,Q) inventory system. For the P_2 service measure only the net stock at the beginning and end of a replenishment cycle are important. If we neglect the correlation between two successive replenishment cycles we may use the formula for the P_2 service measure of the standard (s,Q) inventory system. The average physical stock depends on the sample path of the customer arrivals. Here the assumption of neglecting the correlation structure seems to be more restrictive, since we assume that two successive deliveries are independent. In spite of this, we will see that the formula from the standard (s,Q) model is a very good approximation.

So from the standard (s,Q) model we have (see formulas (2.27) and (2.33) from Chapter 2.3)

$$B_4(s,Q,D_{max},T) \approx \frac{\mathbb{E}((D(L)-s)^+)^2 - \mathbb{E}((D(L)-s-Q)^+)^2}{2Q}, \tag{8.1}$$

$$P_2(s,Q,D_{max},T) \approx 1 - \frac{\mathbb{E}(D(L)+U-s)^+ - \mathbb{E}(D(L)+U-s-Q)^+}{Q}. \tag{8.2}$$

For the optmization we fit ME-distributions to the distribution of $D(L)$ and $D(L)+U$ (see section 2.4 and use local search to find the optimal values for the control variables. What remains to be computed are the first two moments of $D(L)$ and U, and an expression for the number of customer deliveries λ^*.

8.2 The fast approximation method

In this and the following section we describe two approximation methods to determine the moments of $D(L)$ and U for the situation where information about future deliveries is not taken into account explicitly. The difference between this model and the standard (s, Q) model is caused by the method of customer delivery. In the standard (s, Q) model the delivery process and the demand process coincide when the physical stock is sufficiently large. Delivery splitting, however, affects the delivery process in the sense that one single demand now gives rise to a sequence of smaller sized and equally spaced deliveries, even when the physical stock is sufficiently large.

In general it will be very difficult to exactly describe the resulting stochastic process of delivery occurrences. So approximation approaches are necessary. In this section we make the following simplifying assumptions. The stochastic process of delivery occurrences is considered to be a superposition of an (in principal) infinite number of independent compound Poisson processes. The i-th process in this sequence can be considered as the i-th generation demand offspring, i.e. for $i = 0, 1, \ldots$ a delivery occurs at time t in the i-th offspring process if and only if at $t - iT$ an original demand occurred of size larger than iD_{max}. The assumption that the sequence of compound Poisson processes constitutes a sequence of *independent* processes is the simplification that is made here. It is intuitively clear that the validity of this assumption will improve with increasing values of T.

First we introduce some notation, where quantities that refer to the superposition of the offspring processes are indicated with *. For $i = 0, 1, \ldots$ and $j = 1, 2, \ldots$ we define

D_j^*	:=	size of j-th delivery in the superposed process;
$N^*(t)$	:=	total number of deliveries in $(0, t]$;
$D^*(t)$	:=	total amount delivered in $(0, t]$;
U^*	:=	the undershoot under the level s of the superposed delivery process;
λ_i	:=	delivery intensity of the i-th offspring process;
$D_{i,j}$	:=	size of j-th delivery generated by the i-th offspring process;
$N_i(t)$	:=	number of deliveries in $(0, t]$ generated by the i-th offspring process;
$D_i(t)$	:=	total amount delivered in $(0, t]$ due to deliveries generated by i-th offspring process.

Using Walds theorem it can be shown that the following relations hold

$$D^*(t) = \sum_{i=0}^{\infty} D_i(t), \tag{8.3}$$

$$\lambda^* = \sum_{i=0}^{\infty} \lambda_i, \tag{8.4}$$

$$\mathbb{E}D_1^{*n} = \sum_{i=0}^{\infty} \frac{\lambda_i}{\lambda^*} \mathbb{E}D_{i,1}^n \qquad (n = 1, 2, \ldots). \tag{8.5}$$

The appropriate choice for λ_i is given by

$$\lambda_i = \lambda(1 - F_D(iD_{max})) \qquad (i = 0, 1, \ldots). \tag{8.6}$$

Theorem 8.1

$$\mathbb{E}D_1^* = \frac{\lambda}{\lambda^*}\mathbb{E}D \tag{8.7}$$

$$\mathbb{E}D_1^{*2} = \frac{\lambda}{\lambda^*}\Big(\mathbb{E}D^2 - 2D_{max}\sum_{i=1}^{\infty}\int_{iD_{max}}^{\infty}(1-F_D(x))dx\Big) \tag{8.8}$$

See Appendix 8.A.
Note that $\mathbb{E}D_1^{*2} \leq \mathbb{E}D^2$, which indicates a reduction in variance through delivery splitting. In case the distribution function of the demand size is a mixture of Erlang distributions, the third moment can be computed explicitly.

Because L is deterministic and $N^*(t)$ is assumed to be a Poisson process we find for the total amount delivered during the lead time

$$\begin{aligned}
\mathbb{E}D^*(L) &= \mathbb{E}N^*(L)\mathbb{E}D_1^* \\
&= \lambda^* L\, \mathbb{E}D_1^* \\
&= \lambda L\, \mathbb{E}D; && (8.9)\\
\sigma^2(D^*(L)) &= \mathbb{E}N^*(L)\sigma^2(D_1^*) + \sigma^2(N^*(L))\mathbb{E}D_1^{*2} \\
&= \lambda^* L\sigma^2(D_1^*) + \lambda^* L\mathbb{E}D_1^{*2}. && (8.10)
\end{aligned}$$

The distribution of the undershoot has approximately a asymptotic forward recurrence time. Using (2.12) yields

$$\mathbb{E}U^* \approx \frac{\mathbb{E}D_1^{*2}}{2\mathbb{E}D_1^*}, \tag{8.11}$$

$$\mathbb{E}U^{*2} \approx \frac{\mathbb{E}D_1^{*3}}{3\mathbb{E}D_1^*}. \tag{8.12}$$

Combining (8.9) to (8.12) with (8.41) and (8.41) we have expressions for the first two moments of $D^*(L)$ and U^*. Furthermore, we derived an expression for λ^*. It is worthwhile to note that the above described approximation procedure leads to values of s which are independent of the actual value of T. In section 8.4 we evaluate the validity of this approximation procedure.

8.3 A more advanced method

In contrast with the method presented in the previous section we now take the correlation structure between subsequent deliveries into account explicitly. Therefore we need to be more precise about what we mean by the demand during the lead time $D(L)$. Here, $D(L)$ is define as the total amount delivered during the lead time L minus the deliveries due to the particular customer who triggered the replenishment to be delivered within the lead time. U is defined as the undershoot under the reorder point of the delivery that triggers

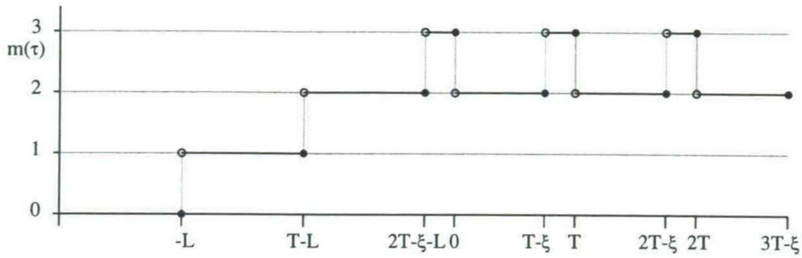


Figure 8.1: The function $m(\tau)$ $(T = 4; L = 9)$

the replenishment plus the remaining deliveries during the lead time caused by the same customer. Hereby are $D(L)$ and U independent.

Assuming that a replenishment order is triggered at time 0 we focus on the delivery process during the interval $[0, L)$ and note that $D(L)$ can be decomposed into two parts:

$D_1(0, L)$:= the total amount delivered during the lead time due to new customers arriving during the lead time,

$D_2(0, L)$:= the total amount delivered during the lead time due to customers who arrived before 0.

To illustrate the influence of various customers on the total amount to be delivered during the lead time L we introduce the function $m(\tau)$ denoting the maximum number of deliveries during $[0, L]$, caused by a customer who arrived at time τ ($\tau \in \mathbb{R}$), when the replenishment was triggered at time 0. In determining an expression for $m(\tau)$ we agree that a replenishment is handled before a delivery when they coincide in time, which occurs with positive probability when L in an integral multiple of T, (see also remark 8.3 at the end of this section). A moment reflection reveals (see Figure 8.1) that

$$m(\tau) = \begin{cases} 0 & \tau \geq L & \\ i & L - iT \leq \tau < L - (i-1)T & i = 1, \ldots, \lfloor \frac{L}{T} \rfloor \\ \lfloor \frac{L}{T} \rfloor & \xi - iT \leq \tau < -(i-1)T & i = 1, 2, \ldots \\ \lfloor \frac{L}{T} \rfloor + 1 & -iT \leq \tau < \xi - iT & i = 0, 1, \ldots \end{cases} \tag{8.13}$$

where $\lfloor x \rfloor := \max\{n \leq x | n \in \mathbb{N}\}$ and $\xi := L - \lfloor \frac{L}{T} \rfloor T$.

Next we define

$N_i^{(1)}$:= number of customers that arrive during the interval $[L-iT, L-(i-1)T)$, $i = 1, 2, \ldots, \lfloor\frac{L}{T}\rfloor$;

$D_{i,j}^{(1)}$:= the contribution to $Z_1(L)$ of the j-th customer arriving in interval $[L-iT, L-(i-1)T)$, $i = 1, \ldots, \lfloor\frac{L}{T}\rfloor$, $j = 1, 2, \ldots, N_i^{(1)}$;

$N_{\lfloor\frac{L}{T}\rfloor+1}^{(1)}$:= number of customers that arrive during the interval $[0, \xi)$;

$D_{\lfloor\frac{L}{T}\rfloor+1,j}^{(1)}$:= the contribution to $Z_1(L)$ of the j-th customer arriving during the interval $[0, \xi)$, $j = 1, 2, \ldots, N_{\lfloor\frac{L}{T}\rfloor+1}^{(1)}$;

$N_{i,A}^{(2)}$:= number of customers that arrive during the interval $[\xi - iT, -(i-1)T)$, $i = 1, 2, \ldots$;

$D_{i,j,A}^{(2)}$:= the contribution to $Z_2(L)$ of the j-th customer arriving during the interval $[\xi - iT, -(i-1)T)$, $i = 1, 2, \ldots$, $j = 1, 2, \ldots, N_{i,A}^{(2)}$;

$N_{i,B}^{(2)}$:= number of customers that arrive during the interval $[-iT, \xi - iT)$, $i = 1, 2, \ldots$;

$D_{i,j,B}^{(2)}$:= the contribution to $Z_2(L)$ of the j-th customer arriving during the interval $[-iT, \xi - iT)$, $i = 1, 2, \ldots$; $j = 1, 2, \ldots, N_{i,B}^{(2)}$.

Then we have

$$D_1(0, L) = \sum_{i=1}^{\lfloor\frac{L}{T}\rfloor+1} \sum_{j=1}^{N_i^{(1)}} D_{i,j}^{(1)} \tag{8.14}$$

$$D_2(0, L) = \sum_{i=1}^{\infty} \left(\sum_{j=1}^{N_{i,A}^{(2)}} D_{i,j,A}^{(2)} + \sum_{j=1}^{N_{i,B}^{(2)}} D_{i,j,B}^{(2)} \right) \tag{8.15}$$

Note that all contributions to $D_1(0, L)$ and $D_2(0, L)$ (that is $D_{i,j}^{(1)}$ and $D_{i,j}^{(2)}$) are mutually independent, because the demands of different customers are independent of each other. Define for a generic random variable D with distribution function F_D,

$$\bar{D}_{k,l} := \min\{(D - kD_{max})^+, lD_{max}\} \qquad (k, l = 0, 1, \ldots). \tag{8.16}$$

Then

$$D_{i,j}^{(1)} \overset{d}{=} \bar{D}_{0,i} \qquad (i = 1, 2, \ldots, \lfloor\frac{L}{T}\rfloor; j = 1, 2, \ldots, N_i^{(1)}), \tag{8.17}$$

and

$$D_{i,j,A}^{(2)} \overset{d}{=} \bar{D}_{i,\lfloor\frac{L}{T}\rfloor} \qquad (i, j = 1, 2, \ldots) \tag{8.18}$$

$$D_{i,j,B}^{(2)} \overset{d}{=} \bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1} \qquad (i, j = 1, 2, \ldots) \tag{8.19}$$

By standard calculus it follows that the n-th moment $(n = 1, 2, \ldots)$ of $\bar{D}_{k,l}$ $(k = 0, 1, \ldots , l = 1, 2, \ldots)$ satisfies

$$\mathbb{E}(\bar{D}_{k,l})^n = \int_{kD_{max}}^{(k+l)D_{max}} (x - kD_{max})^n dF_D(x) + (lD_{max})^n(1 - F_D((k+l)D_{max})). \tag{8.20}$$

Let $y = kD_{max}$ and $x = lD_{max}$ then

$$\begin{aligned}
\mathbb{E}\bar{D}_{k,l} &= \int_{y}^{y+x} (z-y)dF_D(z) + x(1-F_D(y+x)\\
&= \int_{y}^{\infty}(z-y)DF_D(z) - \int_{y+x}^{\infty}(z-y-x)dF_D(z)\\
&= \mathbb{E}(D-kD_{max})^+ - \mathbb{E}(D-(k+l)D_{max})^+,
\end{aligned} \tag{8.21}$$

$$\begin{aligned}
\mathbb{E}\bar{D}^2_{k,l} &= \int_{y}^{y+x} (z-y)^2dF_D(z) + x^2(1-F_D(y+x)\\
&= \int_{y}^{\infty}(z-y)^2DF_D(z) - \int_{y+x}^{\infty}((z-y)^2-x^2)dF_D(z)\\
&= \mathbb{E}((D-kD_{max})^+)^2 - \mathbb{E}((D-(k+l)D_{max})^+)^2\\
&\quad - 2(lD_{max})\mathbb{E}(D-(k+l)D_{max})^+,
\end{aligned} \tag{8.22}$$

and after some algebra

$$\begin{aligned}
\mathbb{E}\bar{D}^3_{k,l} &= \mathbb{E}((D-kD_{max})^+)^3 - \mathbb{E}((D-(k+l)D_{max})^+)^3\\
&\quad - 3(lD_{max})\ \mathbb{E}((D-(k+l)D_{max})^+)^2\\
&\quad - 3(lD_{max})^2\mathbb{E}(D-(k+l)D_{max})^+.
\end{aligned} \tag{8.23}$$

Furthermore, it can be shown that the following relations hold

$$\bar{D}_{k,l} = \bar{D}_{k,1} + \bar{D}_{k+1,l-1} \qquad (k = 0,1,\ldots, l = 1,2,\ldots), \tag{8.24}$$

and

$$\bar{D}_{k,l} = \sum_{j=k}^{k+l-1} \bar{D}_{j,1} \qquad (k = 0,1,\ldots ; l = 1,2,\ldots) \tag{8.25}$$

(for the proof we refer to Appendix 8.B). From (8.14) to (8.19) we conclude that (see appendix 8.C)

$$\begin{aligned}
\mathbb{E}D(L) &= \sum_{i=1}^{\lfloor\frac{L}{T}\rfloor} \lambda T\,\mathbb{E}\bar{D}_{0,i} + (\lambda\xi)\mathbb{E}\bar{D}_{0,\lfloor\frac{L}{T}\rfloor+1}\\
&\quad + \sum_{i=1}^{\infty}\left(\lambda(T-\xi)\mathbb{E}\bar{D}_{i,\lfloor\frac{L}{T}\rfloor} + (\lambda\xi)\mathbb{E}\bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1}\right)\\
&= \lambda L\,\mathbb{E}D
\end{aligned} \tag{8.26}$$

$$\begin{aligned}
\sigma^2(D(L)) &= \sum_{i=1}^{\lfloor\frac{L}{T}\rfloor} \lambda T\,\mathbb{E}\bar{D}^2_{0,i} + (\lambda\xi)\mathbb{E}\bar{D}^2_{0,\lfloor\frac{L}{T}\rfloor+1} \qquad (8.27)\\
&\quad + \sum_{i=1}^{\infty}\left(\lambda(T-\xi)\mathbb{E}(\bar{D}^2_{i,\lfloor\frac{L}{T}\rfloor}) + (\lambda\xi)\mathbb{E}\bar{D}^2_{i,\lfloor\frac{L}{T}\rfloor+1}\right)
\end{aligned}$$

Note that (8.26) is equal to (8.9), but (8.27) differs from (8.10) in the sense that in (8.27) correlations between the offspring processes are taken into account explicitly.

When L is stochastic, expressions for $I\!\!E D(L)$ and $\sigma^2(D(L))$ can be obtained by taking the expectation over (8.26) and (8.27), which yields

$$I\!\!E D(L) = \lambda I\!\!E L\, I\!\!E D \tag{8.28}$$

$$\begin{aligned}\sigma^2(D(L)) &= \sum_{k=0}^{\infty}\sum_{i=1}^{k} \lambda T I\!\!E \bar{D}_{0,i}^2 L_k^0 + \sum_{k=0}^{\infty} \lambda(L_k^1 - kTL_k^0) I\!\!E \bar{D}_{0,k+1}^2 \\ &+ \sum_{k=0}^{\infty}\sum_{i=1}^{\infty}\left(\lambda((k+1)TL_k^0 - L_k^1) I\!\!E(\bar{D}_{i,k}^2) + (\lambda(L_k^1 - kTL_k^0)) I\!\!E \bar{D}_{i,k+1}^2\right)\end{aligned} \tag{8.29}$$

where $L_k^m := \int_{kT}^{(k+1)T} x^m dF_L(x)$. Clearly this expression is very cumbersome to calculate, and that is why we restrict ourselves to deterministic lead times. The reason to use compound Poisson instead of compound renewal is that in the latter case $N_i^{(1)}$, $N_{i,A}^{(2)}$ and $N_{i,A}^{(2)}$ are correlated and accurate moments of them are hard to obtain.

What remains is to obtain expressions for the first two moments of U. For this purpose we follow the approach of section 8.2, where the total delivery process is interpreted as a superposition of independent compound Poisson processes (the so-called offspring processes). (Note that a substantial difference with section 8.2 remains, since here the approach with the offspring processes is only used to compute approximations for U, while in section 8.2 it was used for the computation of the first two moments of both $D(L)$ and U.)
We define for $i = 1, 2, \ldots$:

$\hat{D}_i$:= the remaining amount to be delivered to the customer who triggered a replenishment through the i-th offspring process;

$\hat{D}_{i,k}$:= $\min\{\hat{D}_i, kD_{max}\}$;

U_i := the total contribution to U of the customer who triggered a replenishment through the i-th offspring process;

q_i := the probability that a replenishment is triggered through the i-th offspring process.

Theorem 8.2 $$q_i \approx \frac{\lambda_i I\!\!E D_{i,1}}{\sum_{j=0}^{\infty} \lambda_j I\!\!E D_{j,1}} = \frac{I\!\!E \bar{D}_{i,1}}{I\!\!E D} \quad (i = 0, 1, \ldots),$$

For the proof see Appendix 8.D.

Next we derive an approximation for the probability distribution of $\hat{D}_i$. For this purpose we note that in a standard inventory control process, where demands are generated by a sequence of i.i.d. random variables $(W_j)_{j=1}^{\infty}$, with density function $f_W(x)$, the density function $f_{\hat{W}}(x)$ of the demand $\hat{W}$ that causes the undershoot under the reorder point s can be approximated by (see e.g. Tijms (1994) pp. 85 or Cox (1962) pp. 65-66)

$$f_{\hat{W}}(w) \approx \frac{w f_W(w)}{\mathbb{E}W_1}. \tag{8.30}$$

Denote the inventory position just prior to a replenishment by V. Then it can be shown that V given $\hat{W} = w$ is uniformly distributed over $(s, s+w)$. Now note that in the situation under consideration all actual deliveries (which are also 'trigger'-quantities) are truncated to D_{max}, while the actual value of $\hat{D}_i$ can be larger than D_{max}. Thus realizations of $\hat{D}_i$ bigger than D_{max} have equal probability to trigger, where the probability of triggering a replenishment for realizations of $\hat{D}_i$ smaller than D_{max} are proportional to their actual size (according to (8.30)). Using the appropriate analogue of formula (8.30), the density function of $\hat{D}_i$ can be approximated by

$$f_{\hat{D}_i}(x) \approx \frac{\min\{x, D_{max}\} f_D(x + iD_{max})}{\mathbb{E}\bar{D}_{i,1}} \qquad (i = 0, 1, \ldots). \tag{8.31}$$

Define

V_i := the inventory position just prior to a replenishment when a replenishment is triggered through the i-th off spring process $(i = 0, 1, \ldots)$.

Then using the appropriate analogue of V we conclude that V_i given $\hat{D}_i = d$ is uniformly distributed over $(s, s+\min\{d, D_{max}\})$. Also note that the remaining amount to be delivered within the lead time is at most $\lceil \frac{L}{T} \rceil D_{max}$.
Hence

$$U_i = \hat{D}_{i,\lfloor \frac{L}{T} \rfloor} - V_i \qquad (i = 0, 1, \ldots). \tag{8.32}$$

From (8.32) we find, after some straightforward calculations, the following expressions for the first two moments of U (see Appendix 8.E)

$$\mathbb{E}U \approx \frac{\mathbb{E}D_1^{*2}}{2\mathbb{E}D_1^*} + \frac{D_{max}}{\mathbb{E}D} \sum_{i=0}^{\infty} \mathbb{E}\bar{D}_{(i+1),\lceil \frac{L}{T} \rceil - 1} \tag{8.33}$$

$$\mathbb{E}U^2 \approx \frac{\mathbb{E}D_1^{*3}}{3\mathbb{E}D_1^*} + \frac{D_{max}}{\mathbb{E}D} \sum_{i=0}^{\infty} \mathbb{E}(\bar{D}_{(i+1),\lceil \frac{L}{T} \rceil - 1})^2 + \frac{D_{max}^2}{\mathbb{E}D} \sum_{i=0}^{\infty} \mathbb{E}\bar{D}_{(i+1),\lceil \frac{L}{T} \rceil - 1} \tag{8.34}$$

Remark 8.3 *It can be shown that in case $T = L$ formula (8.27) reduces to (8.10). Moreover, the expressions (8.33) and (8.34) for the first two moments of the undershoot reduce to the corresponding expressions (8.11) and (8.12) in section 8.2. This implies that in case $T = L$ in fact the methods of section 8.2 and 8.3 give the same results.*

Remark 8.4 *In case we agree that a delivery is handled before a replenishment order when they coincide in time, then relation (8.26) and (8.27) still hold. However, the remaining amount to be delivered within the lead time of the customer that triggers the replenishment is at most $\lfloor \frac{L}{T} \rfloor D_{max}$ (instead of $(\lfloor \frac{L}{T} \rfloor - 1)D_{max}$). Thus $U_i = \hat{D}_{i,\lfloor \frac{L}{T} \rfloor + 1} - V_i$ $(i = 0, 1, \ldots)$ and therefore (8.33) and (8.34) have to be adapted accordingly. Note that the method in section 8.2 is invariant for the priority rule for replenishment orders and deliveries.*

8.4 Numerical validation

The previous analysis is an approximation due to two kinds of approximations. Firstly, for the performance measures approximation formulas (8.1) and (8.2) are used. Secondly, when computing these performance measures two approximation methods are developed to compute the first two moments of the demand during the lead time and undershoot. This section validates the quality of the approximations for both methods described in sections 8.2 and 8.3. The results are compared with discrete event simulation. First we computed s by solving $P_2(s, Q, D_{max}, T) = P_{2,target}$ via a numerical search routine. Then, via simulation the actual service levels are determined as well as the average stock on hand for the reorder points generated by the approximation methods. Apart from the usual decision variables (order quantity Q and reorder level s), the supplier has to choose at least values for two other decison variables (the critical demand size D_{max} and the time between two subsequent deliveries T), which makes simulation as such less appropriate as optimization tool.

For each experiment we simulated 5 times 100.000 customers. In this section we considered 120 cases as follows. The average customer demand size is fixed at 50, the coefficient of variation of the customer demand size (c_D) varies among 1,2 and 3 to emphasize the high variability in the demand size. The average number of customer orders (λ) is set equal to 1 per day. The lead time of replenishment orders (L) is equal to 10 days. The replenishment order quantity is taken equal to 1000 for all cases. We adjust the P_2-service level ($P_{2,target}$) as 0.90 and 0.99. The intershipment time T is varied between $\frac{3}{10}$, $\frac{1}{2}$, $\frac{7}{10}$ and 1 times the lead time. The maximum lotsize of a shipment (D_{max}) is varied as 0.5, 1, 2 and 4 times the average customer demand.

Figure 8.2 illustrates the mean absolute deviation of the target service. This shows the poor quality of the 'quick and dirty' method for $T < L$. For $T = L$, however, the 'quick and dirty' method is satisfactory. It can be shown that for $T = L$ both methods give the same results (see Remark 3.1). We notice the good performance of the advanced method for all values of T, as is shown in Figure 8.2. Figure 8.3 illustrates the mean absolute deviation of the average physical stock level. It seems that the 'quick and dirty' method deviates for some cases, but for most cases this method is satisfactory. The advanced method performs good for all values of T. The assumption that deliveries are independent is for the average physical stock not so bad after all. A possible explanation for this good performance is that delivery splitting smoothes the demand process. And it is well known that for the average stock in inventory systems with stable demand processes good and simple approximations are available. Already the simple formula $s + Q/2 - \mathbb{E}D(L)$ is satisfactory in such situations.

The good performance is confirmed by more extensive numerical experiments which are illustrated in Figure 8.4.

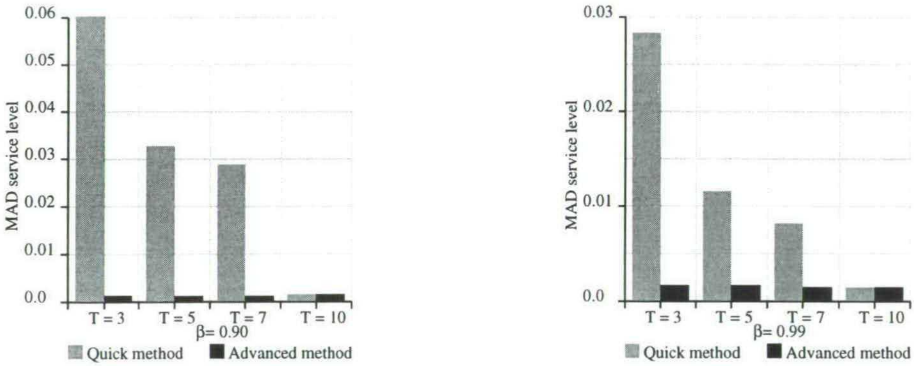


Figure 8.2: The mean absolute deviation of the target service level

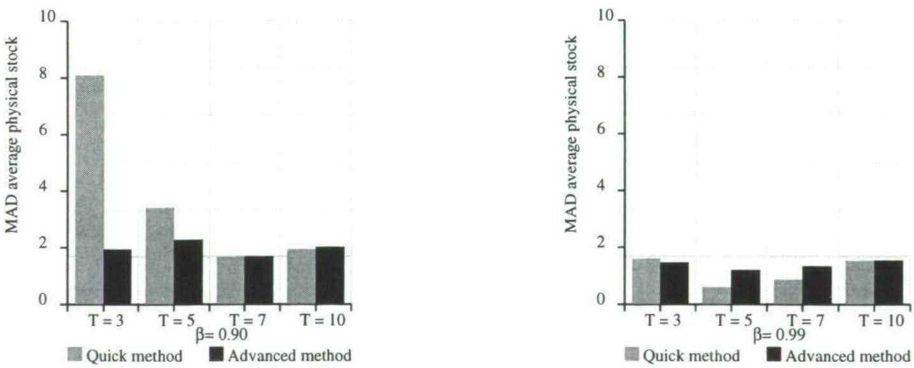


Figure 8.3: The mean absolute deviation of the average physical stock level

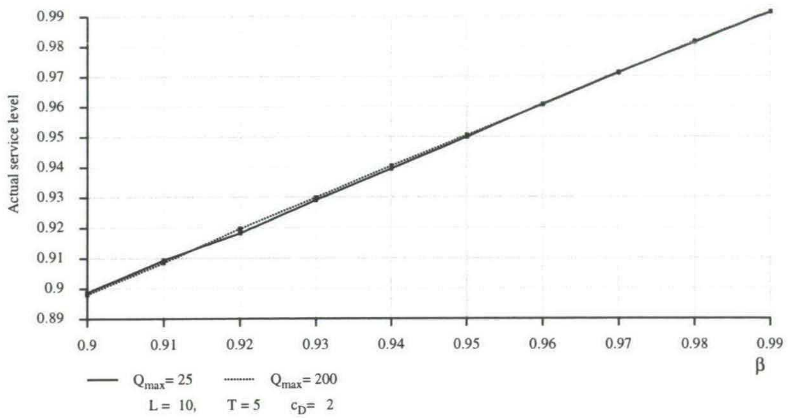


Figure 8.4: The actual service levels associated with the reorder point computed by the more advanced method.

8.5 A replenishment strategy based on known future deliveries

So far we considered a replenishment policy which is only based on the inventory position i.e. physical inventory level plus stock on order minus backorders. However, in case of delivery splitting there exists explicit knowledge about the occurrence of future deliveries of previously splitted orders. This knowledge could be used to improve the performance of the inventory system.

In this section we will deal with an inventory replenishment policy of (s, Q)-type which is not based on the inventory position at time t but on the inventory position at time t minus all planned future deliveries in $(t, t + L]$. This actually resembles the 'available to promise' inventory level as used in MRP-systems, although the 'available to promise' concept in the MRP-context also takes into account the timing of both stock replenishment and customer orders.

We use the following notation:

$$
\begin{aligned}
X(t) &:= \text{inventory position at time } t; \\
K(t) &:= \text{the known deliveries during the interval } (t, t+L]; \\
H(t) &:= X(t) - K(t).
\end{aligned}
$$

In this section the (s, Q) policy prescribes to place a replenishment order of size Q (or multiples of Q) as soon as $H(t)$ drops below the level s. As usual we denote by U, the undershoot under s, the difference between s and $H(t)$ immediately after a replenishment is triggered and $D(L)$ denotes the *unknown* demand during a lead time L. Again formulas

(8.1) and (8.2) can be used to calculate the optimal control parameters. We again resort to the derivation of the first two moments of the independent random variables U and $D(L)$. First we note that $D(L)$ simply equals $D_1(0, L)$ as defined in section 8.3. Hence we conclude from (8.28) and (8.29) that

$$\mathbb{E}(D(L)) = \sum_{i=1}^{\lfloor \frac{L}{T} \rfloor} \lambda T \mathbb{E}(\bar{D}_{0,i}) \tag{8.35}$$

$$\sigma^2(D(L)) = \sum_{i=1}^{\lfloor \frac{L}{T} \rfloor} \lambda T \mathbb{E}(\bar{D}_{0,i}^2) \tag{8.36}$$

To derive the first two moments of U we have to examine the evolution of the process $\{H(t),\ t \geq 0\}$ more closely. Note that $H(t)$ is the difference of two processes $\{X(t),\ t \geq 0\}$ and $\{K(t),\ t \geq 0\}$. Before going into detail, we note that in case L is a integral multiple of T there arises some ambiguity. Consider, for purpose of illustration, the case $T = L$. In this situation it may happen that two events will coincide in time, namely, the arrival of the replenishment order and the actual delivery that triggered the replenishment. When a replenishment order is handled before a delivery in case they coincide in time (as assumed in the previous sections) and the replenishment is triggered by a planned delivery that comes within view at the end of the lead time, then this planned delivery itself does not contribute to the amount delivered during the lead time. For this reason we assume in this section that deliveries are handled before replenishment orders in case both coincide in time. Note that for a new customer there are at most $\lfloor \frac{L}{T} \rfloor + 1$ deliveries within the lead time.

At an arbitrary time epoch t there may occur two possible events which affect the inventory position $X(t)$.

- A new arriving customer at t with demand D decreases $X(t)$ with $\min\{D, D_{max}\}$.
- A customer who arrived at epoch $t - nT\ (n = 1, 2, \ldots)$ with demand $D > nD_{max}$ decreases $X(t)$ with $\min\{D - nD_{max}, D_{max}\}$.

At an arbitrary time epoch t three events may occur which affect $K(t)$

- A new arriving customer at epoch t with demand D increases $K(t)$ with $\min\{D - D_{max}, \lfloor \frac{L}{T} \rfloor D_{max}\}$.
- A customer who arrived at epoch $t - nT\ (n = 1, 2, \ldots)$ with demand $D > nD_{max}$ decreases $K(t)$ with $\min\{D - nD_{max}, D_{max}\}$.
- A customer who arrived at epoch $t + L - nT\ (n = \lfloor \frac{L}{T} \rfloor + 1, \lfloor \frac{L}{T} \rfloor + 2, \ldots)$ with demand $D > nD_{max}$ increases $K(t)$ with $\min\{D - nD_{max}, D_{max}\}$.

Note that the second effect on $K(t)$ is caused by a change from a planned delivery into an actual delivery, while the third effect is caused by a change from a planned delivery outside the lead time period L into a planned delivery inside the lead time period. Also we note that the second effect on $X(t)$ and the second effect on $K(t)$ neutralize each other as far as $H(t)$ is concerned. Hence $H(t)$ is affected by only two events

- A new arriving customer at epoch t with demand D decreases $H(t)$ with $\min\{D, (\lfloor\frac{L}{T}\rfloor + 1)D_{max}\}$ (combining the first effect on $X(t)$ with the first effect on $K(t)$).
- A customer who arrived at epoch $t + L - nT$ ($n = \lfloor\frac{L}{T}\rfloor + 1, \lfloor\frac{L}{T}\rfloor + 2, ...$) with demand $D > nD_{max}$ decreases $H(t)$ with $\min\{D - nD_{max}, D_{max}\}$ (third effect on $K(t)$).

Now we conclude that the amounts by which $H(t)$ decreases (the 'deliveries') are generated by either new customers or by old customers. To approximate the 'delivery' process we again take the viewpoint of section 8.2, where the delivery process is considered as being a superposition of offspring processes. However, note that the jump sizes are different in this situation. Denoting an arbitrary jump size of $H(t)$ by H and following the same line of reasoning as in section 8.2 and using (8.5) for the first moment of H we conclude that (compare formulas (8.4),(8.41) and (8.41))

$$\begin{aligned} \mathbb{E}H &= \frac{\lambda}{\tilde{\lambda}}\mathbb{E}\bar{D}_{0,\lfloor\frac{L}{T}\rfloor+1} + \sum_{i=\lfloor\frac{L}{T}\rfloor+1}^{\infty} \frac{\lambda}{\tilde{\lambda}}\mathbb{E}\bar{D}_{i,1} \\ &= \frac{\lambda}{\tilde{\lambda}}\Big(\sum_{j=0}^{\lfloor\frac{L}{T}\rfloor} \mathbb{E}\bar{D}_{j,1} + \sum_{j=\lfloor\frac{L}{T}\rfloor+1}^{\infty} \mathbb{E}\bar{D}_{j,1}\Big) = \frac{\lambda}{\tilde{\lambda}}\mathbb{E}D, \qquad (8.37) \\ \mathbb{E}H^n &= \frac{\lambda}{\tilde{\lambda}}\Big(\mathbb{E}\bar{D}^n_{0,\lfloor\frac{L}{T}\rfloor+1} + \sum_{i=\lfloor\frac{L}{T}\rfloor+1}^{\infty} \mathbb{E}\bar{D}^n_{i,1}\Big) \qquad (n = 2,3,\ldots), \qquad (8.38) \end{aligned}$$

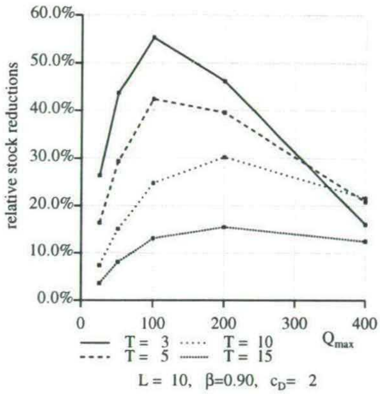
where $\tilde{\lambda} = \lambda + \sum_{i=\lfloor\frac{L}{T}\rfloor+1}^{\infty} \lambda_i$, and λ_i is defined by (8.6). Analogous to (8.11) and (8.12) we conclude

$$\mathbb{E}U \approx \frac{\mathbb{E}H^2}{2\mathbb{E}H}, \qquad (8.39)$$

$$\mathbb{E}U^2 \approx \frac{\mathbb{E}H^3}{3\mathbb{E}H}. \qquad (8.40)$$

Now we again can solve the optimization problem formulated in section 8.1

As for the case in which no information about future deliveries is used, the advanced method also performs very good in case where information about future deliveries is used. To compare the effectiveness of using information about future deliveries explicitly, we adjust the advanced method from section 8.3 for the case a delivery is handled before a replenishment order in case they coincide in time (see Remark 8.4). In Figure 8.5 we present the relative stock reductions obtained by delivery splitting in case we use information about future deliveries over the stock reductions obtained by delivery splitting without using information about future deliveries. To be more precise the figure represents the quantity



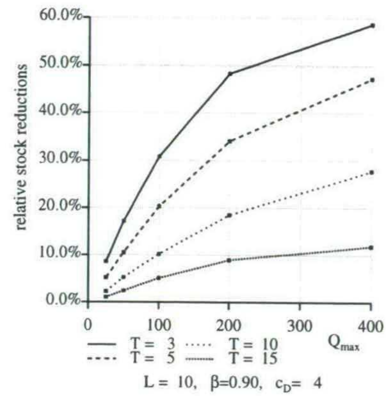


Figure 8.5: Additional stock reductions obtained by using information about future deliveries

$(\mathbb{E}B_4^{ni} - \mathbb{E}B_4^{i})/(\mathbb{E}B_4^{nd} - \mathbb{E}B_4^{ni}) \times 100\%$ where B_4^i denotes the average stock on hand level with delivery splitting using information about future deliveries explicitly, B_4^{ni} denotes the average stock on hand level with delivery splitting without using information about future deliveries, and B_4^{nd} denotes the average stock on hand level without delivery splitting.

It is clear that the additional stock reductions are dependent of D_{max}. Actually, we conjecture that there exists a D_{max} for which the additional stock reductions are maximal. The additional stock reduction increase for small D_{max} because of the increasing amount known to be delivered during the lead time (the number of deliveries within the lead time remains the same but the quantity per delivery increases). Thus the information about future deliveries is used more effectively. On the other hand, for large D_{max} the total amount known to be delivered during the lead time decreases, because the number of future deliveries decreases.

In deciding whether to implement delivery splitting with or without using the information about future deliveries a trade-off has to be made between the additional stock on hand savings and the extra cost due to a more complex replenishment strategy.

8.6 The profitability of delivery splitting

Consider a supply chain consisting of a factory supplying a Regional Distribution Center (RDC), and 10 Local Stockpoints (LSP), from where customer demand is satisfied. The lead time from factory to RDC is deterministic and equals 20 days. The lead times between the RDC and the LSP's are assumed to be independent and identically distributed random variables, with expectation equal to 5 days and with coefficient of variation equal to $\frac{1}{2}$. We assume that all stockpoints have compound renewal demand processes. The expected interarrival time of customers ($\mathbb{E}A$) at LSP 1 to 9 is equal to 5 days, and at LSP 10 equal to 1 day. The coefficient of variation of the interarrival times (c_A) is equal to 1 at LSP 1 to 9, and is $\frac{1}{2}$ at LSP 10. The expected demand size of the customers ($\mathbb{E}D$) at LSP 1 to 9 is equal to 1, and is equal to 5 at LSP 10. The coefficient of variation of the demand size (c_D) for LSP 1 to 9 is equal to $\frac{1}{2}$ and is equal to 1 at LSP 10. The replenishments at the RDC as well as the LSP's are controlled by (s, Q) policies. As customer service measure the P_2 service measure is used. Demands that cannot be delivered directly from shelf are backordered. The aim is to minimize the costs incurred in the supply chain subject to a P_2 service level of 0.95 at the RDC and LSP's. The unit purchase price of the product is equal to \$1000 (e.g. television sets) and the holding costs are equal to 0.20 \$/\$/year, one year is equal to 360 days. The ordering costs at the RDC, which represent setup cost at a factory are equal to \$7000 per replenishment, and at the LSP the ordering costs are equal to \$50 per replenishment. However, at the RDC an "all units discount" is given of \$50 per product when the order is larger than 200. By using the heuristic method, as described in the previous chapter, we can derive the following control variables, see Table 8.1.

Table 8.1: The optimal values of the control values for each of the depots

depot	$(\mathbb{E}A, c_A)$	$(\mathbb{E}D, c_D)$	s	Q	B_4
LSP 1..9	(5,1)	$(1,\frac{1}{2})$	2.93	6	4.88
LSP 10	$(1,\frac{1}{2})$	(5,1)	34.99	200	108.43
RDC	(3.00,0.96)	(20.4,2.47)	375	412	447.13

In Table 8.1, B_4 denotes the expected average physical stock associated to the optimal values of the control values. In this example, indeed the rather stable demand process at the LSP level is transformed in a highly erratic demand process at the RDC (c_D at the RDC is equal to 2.47!). However, in case no discounts where given the optimal value of Q at LSP 10 would be 30, which would result in a more stable demand process at the RDC. Delivery splitting is applied between the RDC and the LSP's 1 to 10. Note that at the LSP's implicitly order splitting (see Chapter 5) is applied. We calculated the reorder level s and the expected average physical stock at the RDC for several values of D_{max} and T using the methods from order splitting, delivery splitting combined through the method proposed in Chapter 7 for solving a two-echelon system (see Table 8.2).

The savings do increase when D_{max} decreases or T increases, which is intuitively clear. However, in view of determining the optimal values of D_{max} and T, we have to take constraints from the buyer into account. The maximal throughput per unit of time between the RDC and the LSP is given by $\frac{D_{max}}{T}$. This maximal throughput has to be larger than the expected demand per unit of time at each of the LSP's, which is given by $\frac{ED}{EA}$. At LSP 10, $\frac{ED}{EA}$ is equal to 5, which excludes, for example, the option $D_{max} = 12$ and $T = 20$. Hence the savings marked with (*) in Table 8.2 are not feasible due to the constraint of LSP 10. This indicates that optimization of D_{max} and T has to be done in consultation with the buyers. Of course for a complete picture of the advantages of delivery splitting,

Table 8.2: The values of s and B_4 at the RDC when delivery splitting is applied

T	D_{max}	s	B_4	savings \$/year
20	100	223	293	30,800
20	50	176	246	40,200(*)
20	25	147	217	46,000(*)
20	12	131	201	49,200(*)
10	100	288	358	17,800
10	50	227	297	30,000
10	25	180	250	39,400(*)
10	12	149	219	45,600(*)
5	100	340	410	7,400
5	50	290	360	17,400
5	25	230	300	29,400
5	12	181	251	39,200(*)

also the increase of transportation costs has to be taken into account as proposed in section 8.2 (see also Hong and Hayya (1992)).

However, in considering the increase of the number of deliveries (and consequently the increase of transportation costs) the following observation is important. In analysing the sales patterns of about 10.000 consumer electronic products in 13 European countries, we found that fast moving products showed rather erratic demand patterns. When delivery splitting is applied for these fast moving products, the shipment frequency remains the same because full truck loads of one product are changed in full truck loads of a number of products for the same customer. Furthermore co-ordinated supply of a number of customers may increase the shipment frequency for each product separately but does not increase the overall shipment frequency. Handling costs, on the other hand, are only slightly affected when shipments consists of complete pallets only (or standard package sizes). However, in case the transportation costs are relevant, the increase in the delivery intensity can be calculated easily as it equals $\lambda^* - \lambda$ (see section 8.2).

In the final experiments we investigate the effects on the optimal control variables s^*, Q^*, and D^*_{max} when solving $(\mathcal{P}_1)$ for a fixed value for T under various values for the system and cost parameters.

We consider the situation in which the cost parameters are $a = 10$ (per replenishment), $b_4 = 0.20$ (on yearly basis), and $k = 0.40$ (per customer delivery), the system parameters are $L = 10$ (days), $T = 4$ (days), $\lambda = 1$ (per day), $\mathbb{E}D = 50$ (units), $c_D = 1$ and $P_{2,target} = 0.90$.

In the first experiment we vary c_D. Then for given value of D_{max} we solved s^* and Q^* from the minimization problem and compute the associate $TRC(s^*, Q^*, D_{max}, T)$, see Figure 8.6. Figure 8.6 shows that the profitability strongly depends on the c_D. For erratic demand processes delivery splitting indeed is profitable. Moreover, we see that D^*_{max} decreases when c_D increases, and that $TRC(s^*, Q^*, D_{max}, T)$ is convex. The last observation enables us to use a simple search routine to find the optimal value of D^*_{max}.

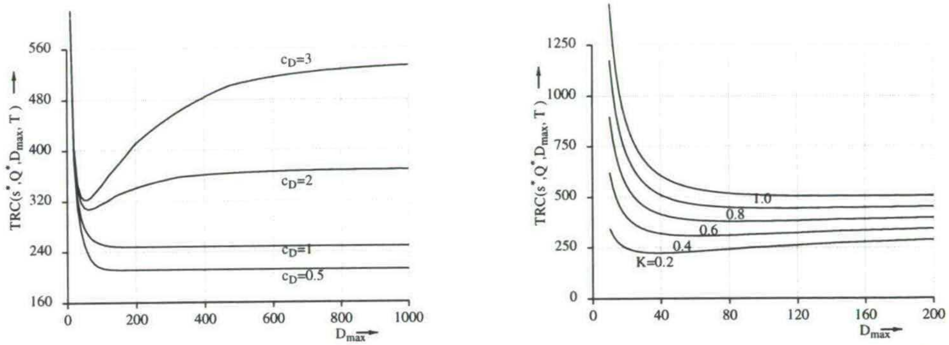


Figure 8.6: The total relevant cost as function of D_{max} for various values of c_D and k.

Now we vary k. Then for given value of D_{max} we solved s^* and Q^* from $(\mathcal{P}_1)$ and computed the associate $C(s^*, Q^*, D_{max}, T)$, see Figure 8.6. Again we see that the profitability depends on the value of k. It is intuitively clear that for high values of the transshipment cost k delivery splitting becomes less attractive.

8.7 Conclusions and future research

Two approximation methods are proposed to calculate moments for the demand during the lead time plus undershoot. The 'quick and dirty' method that is proposed and analysed in section 8.2 only performs satisfactorily when $T = L$. The second more sophisticated method proposed in section 8.3 has an excellent performance irrespective the relation between T and L. The delivery splitting process provides explicit knowledge about future deliveries. This knowledge can be exploited in setting the reorder parameters. In section 8.4 a variant of the method is developed to calculate the appropriate reorder level for the situation in which the information about future deliveries is explicitly used in the replenishment process. Also this method shows excellent performance over a large range of parameter values.

From the numerical examples we conclude that the profitability of delivery splitting strongly depends on the input parameters. Using the approximations as proposed, enables us to find the optimal control variables for the proposed minimization problem.

Several extensions are worthwhile to be considered. The generalisation to stochastic lead times (with the non-overtaking restriction) for the methods described in section 8.2 and 8.3 are straightforward. For the replenishment strategy based on the knowledge of future deliveries a stochastic lead time implies that the total demand during the lead time due to previously splitted orders becomes a stochastic variable, which seems to change the character of the replenishment strategy. Future research will be devoted to this question.

We considered cost minimization under a service level constraint. In order to minimize the sum of inventory, ordering, transhipment and shortage costs we only require an (approximation) expression for the average backlog. It is obvious to use the associated expression from the standard (s, Q) system.

Finally we note that by order splitting not only the stock level of the manufacturer is decreased but also those of the customers. This observation calls for an analysis of the effects of order splitting in a multi-echelon context.

Appendix 8.A: Proof of Theorem 8.1

Expressions for $I\!\!E D_{i,1}$ and $I\!\!E D_{i,1}^2$ are obtained as follows:

$$
\begin{aligned}
& I\!\!E D_{i,1} \\
= \;& I\!\!E((D - iD_{max})I(iD_{max} \leq D < (i+1)D_{max}) + D_{max}I(D \geq (i+1)D_{max})|D > iD_{max}) \\
= \;& \frac{\int_{iD_{max}}^{(i+1)D_{max}} (x - iD_{max})dF_D(x) + \int_{(i+1)D_{max}}^{\infty} D_{max}dF_D(x)}{I\!\!P(D > iD_{max})} \\
= \;& \frac{\int_{iD_{max}}^{(i+1)D_{max}} xdF_D(x) - iD_{max} \int_{iD_{max}}^{(i+1)D_{max}} dF_D(x) + D_{max}(1 - F_D((i+1)D_{max})}{1 - F_D(iD_{max})} \\
= \;& \frac{\int_{iD_{max}}^{(i+1)D_{max}} xdF_D(x) + iD_{max}\Big(F_D(iD_{max}) - F_D((i+1)D_{max})\Big) + D_{max}(1-F_D((i+1)D_{max}))}{1 - F_D(iD_{max})} \\
= \;& \frac{\int_{iD_{max}}^{(i+1)D_{max}} xdF_D(x) + Y_{i+1} - Y_i}{1 - F_D(iD_{max})}
\end{aligned}
$$

with $Y_i = iD_{max}(1 - F_D(iD_{max}))$.
Analogously

$$
\begin{aligned}
& I\!\!E D_{i,1}^2 \\
= \;& I\!\!E((D - iD_{max})^2 I\{iD_{max} \leq D < (i+1)D_{max}\} + D_{max}^2 I\{D \geq (i+1)D_{max}\}|D \geq iD_{max}) \\
= \;& \frac{\int_{iD_{max}}^{(i+1)D_{max}} (x - iD_{max})^2 dF_D(x) + \int_{(i+1)D_{max}}^{\infty} D_{max}^2 dF_D(x)}{1 - F_D(iD_{max})} \\
= \;& \frac{W_i - 2iD_{max}^2(1 - F_D(iD_{max})) + (iD_{max})^2(1 - F_D(iD_{max}))}{1 - F_D(iD_{max})} \\
& - \frac{(i^2-1)D_{max}^2(1 - F_D((i+1)D_{max}))}{1 - F_D(iD_{max})} \\
= \;& \frac{W_i + i(i-2)D_{max}^2\Big(1 - F_D(iD_{max})\Big) - (i+1)(i+1-2)D_{max}^2\Big(1- F_D((i+1)D_{max})\Big)}{1 - F_D(iD_{max})} \\
= \;& \frac{W_i + Z_i - Z_{i+1}}{1 - F_D(iD_{max})}
\end{aligned}
$$

with

$$
W_i = \int_{iD_{max}}^{(i+1)D_{max}} (x^2 - 2iD_{max}x)dF_D(x) + 2iD_{max}^2(1 - F_D(iD_{max}))
$$

and

$$Z_i = (i(i-2)D_{max}^2)(1 - F_D(iD_{max})).$$

Subsitution of the $\mathbb{E}D_{i,1}$ and $\mathbb{E}D_{i,1}^2$ for $i = 0, 1, ...$ in the expressions for $\mathbb{E}D_1^*$ and $\mathbb{E}D_1^{*2}$ and using (8.5) yields

$$\begin{aligned}
\mathbb{E}D_1^* &= \sum_{i=0}^{\infty} \frac{\lambda_i}{\lambda^*} \mathbb{E}D_{i,1} \\
&= \sum_{i=0}^{\infty} \frac{\lambda}{\lambda^*} \Big(\int_{iD_{max}}^{(i+1)D_{max}} x dF_D(x) - Y_i + Y_{i+1} \Big) \\
&= \frac{\lambda}{\lambda^*} \mathbb{E}D
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}D_1^{*2} &= \sum_{i=0}^{\infty} \frac{\lambda_i}{\lambda^*} \mathbb{E}D_{i,1}^2 \\
&= \frac{\lambda}{\lambda^*} \sum_{i=0}^{\infty} \Big(\int_{iD_{max}}^{(i+1)D_{max}} x^2 dF_D(x) - 2D_{max}\big(i \int_{iD_{max}}^{(i+1)D_{max}} x dF_D(x) - \int_{iD_{max}}^{\infty} iD_{max} dF_D(x)\big) + Z_i - Z_{i+1} \Big) \\
&= \frac{\lambda}{\lambda^*} \Big(\mathbb{E}D^2 - 2D_{max} \sum_{i=1}^{\infty} \int_{iD_{max}}^{\infty} (x - iD_{max}) dF_D(x) \Big) \\
&= \frac{\lambda}{\lambda^*} \Big(\mathbb{E}D^2 - 2D_{max} \sum_{i=1}^{\infty} \int_{iD_{max}}^{\infty} (1 - F_D(x)) dx \Big)
\end{aligned}$$

$\square$

Appendix 8.B: Proof of equations (8.24) and (8.25)

Before we proof (8.25) we first proof (8.24). For $l = 1$ it easily follows that $\bar{D}_{k,1} = \bar{D}_{k,1} + \bar{D}_{k+1,0} = \bar{D}_{k,1}$.
For $l > 1$ we distinguish four cases

$$\begin{array}{lll}
D < kD_{max} & : & \bar{D}_{k,l} = 0 \\
 & : & \bar{D}_{k,1} = \bar{D}_{k+1,l-1} = 0 \\
kD_{max} \leq D < (k+1)D_{max} & : & \bar{D}_{k,l} = D - kD_{max} \\
 & : & \bar{D}_{k,1} = D - kD_{max} \\
 & : & \bar{D}_{k+1,l-1} = 0 \\
(k+1)D_{max} \leq D < lD_{max} & : & \bar{D}_{k,l} = D - kD_{max} \\
 & : & \bar{D}_{k,1} = D_{max} \\
 & : & \bar{D}_{k+1,l-1} = D - (k+1)D_{max} \\
D \geq lD_{max} & : & \bar{D}_{k,l} = lD_{max} \\
 & : & \bar{D}_{k,1} = D_{max} \\
 & : & \bar{D}_{k+1,l-1} = (l-1)D_{max}
\end{array}$$

Using (8.24) l times yield to

$$\bar{D}_{k,l} = \sum_{j=k}^{k+l-1} \bar{D}_{j,1} \quad k, l = 0, 1, 2, ... \tag{8.41}$$

which completes the proof of (8.25) □

Appendix 8.C: Proof of relations (8.26) and (8.27)

Let $N := (N_i^{(1)}, i = 1,2,...,\lfloor \frac{L}{T} \rfloor + 1;\ N_{i,A}^{(2)}, i = 1,2,...;\ N_{i,B}^{(2)}, i = 1,2,...)$ Then

$$I\!E(D(L)|N) = I\!E\Big(\sum_{i=1}^{\lfloor \frac{L}{T} \rfloor+1} \sum_{j=1}^{N_i^{(1)}} D_{i,j}^{(1)} + \sum_{i=1}^{\infty} \Big(\sum_{j=1}^{N_{i,A}^{(2)}} D_{i,j,A}^{(2)} + \sum_{j=1}^{N_{i,B}^{(2)}} D_{i,j,B}^{(2)} \Big) \Big| N \Big)$$

Using (8.14) to (8.25) we conclude

$$I\!E(D(L)|N) = \sum_{i=1}^{\lfloor \frac{L}{T} \rfloor+1} N_i^{(1)} I\!E\bar{D}_{0,i} + \sum_{i=1}^{\infty} \Big(N_{i,A}^{(2)} I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor} + N_{i,B}^{(2)} I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor+1} \Big)$$

$$\begin{aligned} I\!ED(L) &= I\!E(I\!E(D(L)|N)) \\ &= I\!E\Big(\sum_{i=1}^{\lfloor \frac{L}{T} \rfloor+1} N_i^{(1)} I\!E\bar{D}_{0,i} + \sum_{i=1}^{\infty} \Big(N_{i,A}^{(2)} I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor} + N_{i,B}^{(2)} I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor+1} \Big) \Big) \\ &= \sum_{i=1}^{\lfloor \frac{L}{T} \rfloor+1} I\!EN_i^{(1)} I\!E\bar{D}_{0,i} + \sum_{i=1}^{\infty} \Big(I\!EN_{i,A}^{(2)} I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor} + I\!EN_{i,B}^{(2)} I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor+1} \Big) \end{aligned}$$

Since the customer arrival process is a Poisson process it follows that

$$\begin{array}{llllll} I\!EN_i^{(1)} & = & \sigma^2(N_i^{(1)}) & = & \lambda T & i = 1,2,...,\lfloor \frac{L}{T} \rfloor \\ I\!EN_{\lfloor \frac{L}{T} \rfloor+1}^{(1)} & = & \sigma^2(N_{\lfloor \frac{L}{T} \rfloor+1}^{(1)}) & = & \lambda\xi & \\ I\!EN_{i,A}^{(2)} & = & \sigma^2(N_{i,A}^{(2)}) & = & \lambda(T-\xi) & i = 1,2,... \\ I\!EN_{i,B}^{(2)} & = & \sigma^2(N_{i,B}^{(2)}) & = & \lambda\xi & i = 1,2,... \end{array}$$

Hence

$$\begin{aligned} I\!ED(L) &= \sum_{i=1}^{\lfloor \frac{L}{T} \rfloor} \lambda T I\!E\bar{D}_{0,i} + \lambda\xi I\!E\bar{D}_{0,\lfloor \frac{L}{T} \rfloor+1} \\ &\quad + \sum_{i=1}^{\infty} \Big(\lambda(T-\xi) I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor} + \lambda\xi I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor+1} \Big), \end{aligned}$$

which yields

$$\begin{aligned} I\!ED(L) &= \sum_{i=1}^{\lfloor \frac{L}{T} \rfloor} \lambda T I\!E\bar{D}_{0,i} + \lambda\xi I\!E\bar{D}_{0,\lfloor \frac{L}{T} \rfloor+1} + \sum_{i=1}^{\infty} \Big(\lambda(T-\xi) I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor} + \lambda\xi I\!E\bar{D}_{i,\lfloor \frac{L}{T} \rfloor+1} \Big) \\ &= \sum_{i=1}^{\lfloor \frac{L}{T} \rfloor} \sum_{j=0}^{i-1} \lambda T I\!E\bar{D}_{j,1} + \sum_{j=0}^{\lfloor \frac{L}{T} \rfloor} \lambda\xi I\!E\bar{D}_{j,1} \\ &\quad + \sum_{i=1}^{\infty} \Big(\lambda(T-\xi) I\!E \sum_{j=i}^{i+\lfloor \frac{L}{T} \rfloor-1} \bar{D}_{j,1} + \lambda\xi \sum_{j=i}^{i+\lfloor \frac{L}{T} \rfloor} I\!E\bar{D}_{j,1} \Big) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=0}^{\lfloor\frac{L}{T}\rfloor-1}\sum_{i=j+1}^{\lfloor\frac{L}{T}\rfloor}\lambda T\mathbb{E}\bar{D}_{j,1}+\sum_{j=0}^{\lfloor\frac{L}{T}\rfloor}\lambda\xi\mathbb{E}\bar{D}_{j,1}\\
&\quad+\sum_{i=1}^{\infty}\sum_{j=i}^{i+\lfloor\frac{L}{T}\rfloor-1}\lambda T\mathbb{E}\bar{D}_{j,1}+\sum_{i=1}^{\infty}\lambda\xi\mathbb{E}\bar{D}_{i+\lfloor\frac{L}{T}\rfloor,1}\\
&= \sum_{j=0}^{\lfloor\frac{L}{T}\rfloor-1}\sum_{i=j+1}^{\lfloor\frac{L}{T}\rfloor}\lambda T\mathbb{E}\bar{D}_{j,1}+\sum_{j=0}^{\infty}\lambda\xi\mathbb{E}\bar{D}_{j,1}\\
&\quad+\sum_{j=0}^{\lfloor\frac{L}{T}\rfloor-1}\sum_{i=1}^{j}\lambda T\mathbb{E}\bar{D}_{j,1}+\sum_{j=\lfloor\frac{L}{T}\rfloor}^{\infty}\sum_{i=j-\lfloor\frac{L}{T}\rfloor+1}^{j}\lambda T\mathbb{E}\bar{D}_{j,1}\\
&= \sum_{j=0}^{\infty}\lambda(\lfloor\frac{L}{T}\rfloor T+\xi)\bar{D}_{j,1}\\
&= \lambda L\mathbb{E}D.
\end{aligned}
$$

For $\sigma^2(D(L))$ we derive

$$
\sigma^2(D(L)|N=n) \;=\; \sigma^2\Bigg(\sum_{i=1}^{\lfloor\frac{L}{T}\rfloor+1}\sum_{j=1}^{N_i^{(1)}}D_{i,j}^{(1)}+\sum_{i=1}^{\infty}\Bigg(\sum_{j=1}^{N_{i,A}^{(2)}}D_{i,j,A}^{(2)}+\sum_{j=1}^{N_{i,B}^{(2)}}D_{i,j,B}^{(2)}\Bigg)\Bigg|N\Bigg).
$$

Hence it follows from (8.14) to (8.25) that

$$
\sigma^2(D(L)|N) \;=\; \sum_{i=1}^{\lfloor\frac{L}{T}\rfloor+1}N_i^{(1)}\sigma^2(\bar{D}_{0,i})+\sum_{i=1}^{\infty}\Big(N_{i,A}^{(2)}\sigma^2(\bar{D}_{i,\lfloor\frac{L}{T}\rfloor})+N_{i,B}^{(2)}\sigma^2(\bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1})\Big),
$$

which implies that

$$
\begin{aligned}
\sigma^2(D(L)) &= \mathbb{E}(\sigma^2(D_L^*|N))+\sigma^2(\mathbb{E}(D_L^*|N))\\
&= \mathbb{E}\Bigg(\Bigg(\sum_{i=1}^{\lfloor\frac{L}{T}\rfloor+1}N_i^{(1)}\sigma^2(\bar{D}_{0,i})+\sum_{i=1}^{\infty}\Big(N_{i,A}^{(2)}\sigma^2(\bar{D}_{i,\lfloor\frac{L}{T}\rfloor})+N_{i,B}^{(2)}\sigma^2(\bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1})\Big)\Bigg)\\
&\quad+\sigma^2\Bigg(\sum_{i=1}^{\lfloor\frac{L}{T}\rfloor+1}N_i^{(1)}\mathbb{E}\bar{D}_{0,i}+\sum_{i=1}^{\infty}\Big(N_{i,A}^{(2)}\mathbb{E}\bar{D}_{i,\lfloor\frac{L}{T}\rfloor}+N_{i,B}^{(2)}\mathbb{E}\bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1}\Big)\Bigg)\\
&= \sum_{i=1}^{\lfloor\frac{L}{T}\rfloor+1}\mathbb{E}N_i^{(1)}\sigma^2(\bar{D}_{0,i})+\sum_{i=1}^{\infty}\Big(\mathbb{E}N_{i,A}^{(2)}\sigma^2(\bar{D}_{i,\lfloor\frac{L}{T}\rfloor})+\mathbb{E}N_{i,B}^{(2)}\sigma^2(\bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1})\Big)\\
&\quad+\sum_{i=1}^{\lfloor\frac{L}{T}\rfloor+1}\sigma^2(N_i^{(1)})\mathbb{E}^2\bar{D}_{0,i}+\sum_{i=1}^{\infty}\Big(\sigma^2(N_{i,A}^{(2)})\mathbb{E}^2\bar{D}_{i,\lfloor\frac{L}{T}\rfloor}+\sigma^2(N_{i,B}^{(2)})\mathbb{E}^2\bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1}\Big)\\
&= \sum_{i=1}^{\lfloor\frac{L}{T}\rfloor}\lambda T\mathbb{E}\bar{D}_{0,i}^2+\lambda\xi\mathbb{E}\bar{D}_{0,\lfloor\frac{L}{T}\rfloor+1}^2+\sum_{i=1}^{\infty}\Big(\lambda(T-\xi)\mathbb{E}\bar{D}_{i,\lfloor\frac{L}{T}\rfloor}^2+\lambda\xi\mathbb{E}\bar{D}_{i,\lfloor\frac{L}{T}\rfloor+1}^2\Big).
\end{aligned}
$$

□

Appendix 8.D: Proof of Theorem 8.2

Let $Y_i(s)$, $i = 0, 1, 2, \ldots$ be a compound Poisson process with Poisson parameter λ_i and demand size distribution $F_i(.)$ with mean $\mathbb{E}D_i$ and $F_i(0) = 0$. Assume that the processes $(Y_i(s))_{i=0}^{\infty})$ are independent and define $X(s) := \sum_{i=0}^{\infty} Y_i(s)$. We are interested in the probability

$q_i(t) :=$ probability that overshoot of $X(s)$ over the value t is caused by the i-th process. Then $q_i(t)$ satisfies the following renewal type equation

$$q_i(t) = \frac{\lambda_i}{\lambda^*}(1 - F_i(t)) + \sum_{j=0}^{\infty} \frac{\lambda_j}{\lambda^*} \int_0^t q_i(t-x)dF_j(x) \tag{8.42}$$

Taking Laplace transforms on both sides yields:

$$\tilde{q}_i(s) = \frac{\lambda_i}{\lambda^* s}(1 - \tilde{F}_i(s)) + \sum_{j=0}^{\infty} \frac{\lambda_j \tilde{q}_i(s)}{\lambda^*} \tilde{F}_j(s) \tag{8.43}$$

where $\tilde{q}_i(s) := \int_0^\infty e^{-st} q_i(t)dt$ and $\tilde{F}_j(s) := \int_0^\infty e^{-st} dF_j(t)$
Solving (8.42) for $\tilde{q}_i(s)$ yields

$$\tilde{q}_i(s) = \frac{\frac{\lambda_i}{\lambda^*}(1 - \tilde{F}_i(s))}{1 - \sum_{j=0}^{\infty} \frac{\lambda_j}{\lambda^*} \tilde{F}_j(s)} \tag{8.44}$$

Since

$$\lim_{t\to\infty} q_i(t) = \lim_{s\downarrow 0} s\tilde{q}_i(s) \tag{8.45}$$

we conclude from (8.43) that

$$\lim_{t\to\infty} q_i(t) = \lim_{s\downarrow 0} \frac{\frac{\lambda_i}{\lambda^*}(1 - \tilde{F}_i(s)}{1 - \sum_{j=0}^{\infty} \frac{\lambda_j}{\lambda^*} \tilde{F}_j(s)} \tag{8.46}$$

which implies, using l'Hopital's rule

$$\lim_{t\to\infty} q_i(t) = \frac{\lambda_i \mathbb{E}(D_i)}{\sum_{j=0}^{\infty} \lambda_j \mathbb{E}(D_j)}, \tag{8.47}$$

which proves the asymptotic validity of (8.2) □

Appendix 8.E: Proof of equations (8.33) and (8.34)

Using (8.2),(8.31) and (8.32) we find

$$
\begin{aligned}
I\!E U_i &= I\!E(\hat{D}_{i,\lceil \frac{L}{T} \rceil} - V_i) \\
&= \int_0^{\lceil \frac{L}{T} \rceil D_{max}} I\!E(x - V_i) f_{\hat{D}_i}(x)dx + \int_{\lceil \frac{L}{T} \rceil D_{max}}^{\infty} I\!E(\lceil \frac{L}{T} \rceil D_{max} - V_i) f_{\hat{D}_i}(x)dx \\
&= \int_0^{D_{max}} \int_0^{x} \frac{x - v}{x} dv f_{\hat{D}_i}(x)dx + \int_{D_{max}}^{\lceil \frac{L}{T} \rceil D_{max}} \int_0^{D_{max}} \frac{x - v}{D_{max}} dv f_{\hat{D}_i}(x)dx \\
&\quad + \int_{\lceil \frac{L}{T} \rceil D_{max}}^{\infty} \int_0^{D_{max}} \frac{\lceil \frac{L}{T} \rceil D_{max} - v}{D_{max}} dv f_{\hat{D}_i}(x)dx \\
&= \int_0^{D_{max}} \frac{1}{2} x f_{\hat{D}_i}(x)dx + \int_{D_{max}}^{\lceil \frac{L}{T} \rceil D_{max}} \Big((x - D_{max}) + \frac{1}{2} D_{max} \Big) f_{\hat{D}_i}(x)dx \\
&\quad + \int_{\lceil \frac{L}{T} \rceil D_{max}}^{\infty} (\lceil \frac{L}{T} \rceil - 1 + \frac{1}{2}) D_{max} f_{\hat{D}_i}(x)dx \\
&= I\!E(\bar{D}_{i,1})^{-1} \Bigg(\int_0^{D_{max}} \frac{1}{2} x^2 f_D(x + iD_{max})dx + \frac{1}{2} D_{max}^2 \int_{D_{max}}^{\infty} f_D(x + iD_{max})dx \\
&\quad + \int_{D_{max}}^{\lceil \frac{L}{T} \rceil D_{max}} D_{max}(x - D_{max}) f_D(x + iD_{max})dx + \int_{\lceil \frac{L}{T} \rceil D_{max}}^{\infty} (\lceil \frac{L}{T} \rceil - 1) D_{max}^2 f_D(x + iD_{max})dx \Bigg) \\
&= I\!E(\bar{D}_{i,1})^{-1} \Bigg(\int_{iD_{max}}^{(i+1)D_{max}} \frac{1}{2} (x - iD_{max})^2 f_D(x)dx + \frac{1}{2} D_{max}^2 \int_{(i+1)D_{max}}^{\infty} f_D(x)dx \\
&\quad + \int_{(i+1)D_{max}}^{(\lceil \frac{L}{T} \rceil + i)D_{max}} D_{max}(x - (i+1)D_{max}) f_D(x)dx + \int_{(\lceil \frac{L}{T} \rceil + i)D_{max}}^{\infty} (\lceil \frac{L}{T} \rceil - 1) D_{max}^2 f_D(x)dx \Bigg) \\
&= \frac{\frac{1}{2} I\!E \bar{D}_{i,1}^2 + D_{max} I\!E \bar{D}_{i+1,\lceil \frac{L}{T} \rceil - 1}}{I\!E \bar{D}_{i,1}}.
\end{aligned}
$$

Analogously

$$
\begin{aligned}
\mathbb{E}U_i^2 &= \mathbb{E}((\hat{D}_{i,\lceil\frac{L}{T}\rceil} - V_i)^2) \\
&= \int_0^{\lceil\frac{L}{T}\rceil D_{max}} \mathbb{E}((x-V_i)^2) f_{\hat{D}_i}(x)dx + \int_{\lceil\frac{L}{T}\rceil D_{max}}^{\infty} \mathbb{E}(((\lceil\frac{L}{T}\rceil D_{max} - V_i)^2) f_{\hat{D}_i}(x)dx \\
&= \int_0^{D_{max}} \int_0^{x} \frac{(x-v)^2}{x} dv f_{\hat{D}_i}(x)dx + \int_{D_{max}}^{\lceil\frac{L}{T}\rceil D_{max}} \int_0^{D_{max}} \frac{(x-v)^2}{D_{max}} dv f_{\hat{D}_i}(x)dx \\
&\quad + \int_{\lceil\frac{L}{T}\rceil D_{max}}^{\infty} \int_0^{D_{max}} \frac{(\lceil\frac{L}{T}\rceil D_{max} - v)^2}{D_{max}} dv f_{\hat{D}_i}(x)dx \\
&= \int_0^{D_{max}} \frac{1}{3}x^2 f_{\hat{D}_i}(x)dx + \int_{D_{max}}^{\lceil\frac{L}{T}\rceil D_{max}} \left(x^2 - xD_{max} + \frac{1}{3}D_{max}^2\right) f_{\hat{D}_i}(x)dx \\
&\quad + \int_{\lceil\frac{L}{T}\rceil D_{max}}^{\infty} ((\lceil\frac{L}{T}\rceil - 1)^2 + \lceil\frac{L}{T}\rceil - 1 + \frac{1}{3}) D_{max}^2 f_{\hat{D}_i}(x)dx \\
&= \mathbb{E}(\bar{D}_{i,1})^{-1}\Bigg(\int_0^{D_{max}} \frac{1}{3}x^3 f_D(x+iD_{max})dx + \frac{1}{3}D_{max}^3 \int_{D_{max}}^{\infty} f_D(x+iD_{max})dx \\
&\quad + \int_{D_{max}}^{\lceil\frac{L}{T}\rceil D_{max}} D_{max}x(x-D_{max}) f_D(x+iD_{max})dx \\
&\quad + \int_{\lceil\frac{L}{T}\rceil D_{max}}^{\infty} ((\lceil\frac{L}{T}\rceil - 1)^2 + \lceil\frac{L}{T}\rceil - 1) D_{max}^3 f_D(x+iD_{max})dx \Bigg) \\
&= \mathbb{E}(\bar{D}_{i,1})^{-1}\Bigg(\frac{1}{3}\mathbb{E}(\bar{D}_{i,1}^3) + D_{max}\Bigg(\int_{(i+1)D_{max}}^{(\lceil\frac{L}{T}\rceil + i)D_{max}} (x-(i+1)D_{max})^2 f_D(x)dx \\
&\quad + \int_{(\lceil\frac{L}{T}\rceil + i)D_{max}}^{\infty} ((\lceil\frac{L}{T}\rceil - 1)D_{max})^2 f_D(x)dx \Bigg) \\
&\quad + D_{max}^2\Bigg(\int_{(i+1)D_{max}}^{(\lceil\frac{L}{T}\rceil + i)D_{max}} (x-(i+1)D_{max}) f_D(x)dx + \int_{(\lceil\frac{L}{T}\rceil + i)D_{max}}^{\infty} (\lceil\frac{L}{T}\rceil - 1)D_{max} f_D(x)dx \Bigg)\Bigg) \\
&= \frac{\frac{1}{3}\mathbb{E}\bar{D}_{i,1}^3 + D_{max}\mathbb{E}\bar{D}_{i+1,\lceil\frac{L}{T}\rceil - 1}^2 + D_{max}^2 \mathbb{E}\bar{D}_{i+1,\lceil\frac{L}{T}\rceil - 1}}{\mathbb{E}\bar{D}_{i,1}}
\end{aligned}
$$

Thus

$$\begin{aligned}
\mathbb{E}U &= \sum_{i=0}^{\infty} q_i \mathbb{E}U_i \\
&= \sum_{i=0}^{\infty} \frac{\mathbb{E}\bar{D}_{i,1}}{\mathbb{E}D}\left(\frac{\mathbb{E}\bar{D}_{i,1}^2}{2\mathbb{E}\bar{D}_{i,1}} + \frac{D_{max}\mathbb{E}\bar{D}_{i+1,\lceil \frac{L}{T}\rceil -1}}{\mathbb{E}\bar{D}_{i,1}}\right) \\
&= \sum_{i=0}^{\infty}\left(\frac{\lambda_i \mathbb{E}D_{i,1}^2}{\lambda^* 2\mathbb{E}D^*} + \frac{D_{max}\mathbb{E}\bar{D}_{i+1,\lceil \frac{L}{T}\rceil -1}}{\mathbb{E}D}\right) \\
&= \frac{\mathbb{E}D^{*2}}{2\mathbb{E}D^*} + \sum_{i=0}^{\infty}\frac{D_{max}\mathbb{E}\bar{D}_{i+1,\lceil \frac{L}{T}\rceil -1}}{\mathbb{E}D}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}U^2 &= \sum_{i=0}^{\infty} q_i \mathbb{E}U_i^2 \\
&= \sum_{i=0}^{\infty} \frac{\mathbb{E}\bar{D}_{i,1}}{\mathbb{E}D}\left(\frac{\mathbb{E}\bar{D}_{i,1}^3}{\mathbb{E}3\bar{D}_{i,1}} + \frac{D_{max}\mathbb{E}\bar{D}_{i+1,\lceil \frac{L}{T}\rceil -1}^2 + D_{max}^2\mathbb{E}\bar{D}_{i+1,\lceil \frac{L}{T}\rceil -1}}{\mathbb{E}\bar{D}_{i,1}}\right) \\
&= \frac{\mathbb{E}D^{*3}}{3\mathbb{E}D^*} + \sum_{i=0}^{\infty}\frac{D_{max}\left(\mathbb{E}\bar{D}_{i+1,\lceil \frac{L}{T}\rceil -1}^2 + D_{max}\mathbb{E}\bar{D}_{i+1,\lceil \frac{L}{T}\rceil -1}\right)}{\mathbb{E}D}
\end{aligned}$$

□

References

AARDAL, K., O. JONSSON, AND H. JONSSON (1989). Optimal inventory policies with service-level constraints. *Journal of Operational Research Society*, **40**, 65-73.

ACKOFF, R.L. (1967). Management misinformation systems. *Management science*, **14**, 147-156.

ADAN, I.J.B.F., M.J.A. EENIGE, AND J.A.C. RESING (1995). Fitting discrete distributions on the first two moments. *Probability in the Engineering and Informational Sciences*, **9**, 623-632.

AXSÄTER, S. (1993). Continuous review policies for multi-level inventory systems with stochastic demand, in Graves, S.C., A.H.G. Rinnooy Kan and P.H. Zipkin (eds.), *Logistics of production and inventory*, Handbooks in operations research and management science 4, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, Chapter 4, 175-197.

BALAKRISHNAN, N. (1988). Recurrence relations for order statistics from n independent and non-identically distributed random variables. *Annals of Institutional Statistics Mathematics*, **40**, 273-277.

BALAKRISHNAN, N., AND A.C. COHEN (1991). *Order statistics and inference*, Academic Press, Inc., San Diego.

BURGIN, T.A. (1975). The gamma distribution and inventory control. *Operational Research Quarterly*, **26**, 507-525.

CHAUDRY, M. L. (1992). Computing stationary queueing-time distributions of $GI/D/1$ and $GI/D/c$ queues. *Naval Research Logistics*, **39**, 975-996.

CHEN, F., AND Y.S. ZHENG (1993). Inventory models with general backorder costs. *European Journal of Operational Research*, **65**, 175-186.

CHIANG, C., AND W.C. BENTON (1994). Sole sourcing versus dual sourcing under stochastic demands and lead times. *Naval Research Logistics*, **41**, 609-624.

CHIANG, C., AND W.C. CHIANG (1996). Reducing inventory costs by order splitting in the sole sourcing environment. *Journal of Operational Research Society*, **47**, 446-456.

CINLAR, E. (1975). *Introduction to stochastic processes*, Englewood Cliffs, Prentice-Hall, London.

COX, D.R. (1962). *Renewal theory*, Methuen, London.

DAS, C. (1985). Approximate solution to the (Q, r) inventory model for gamma lead time demand. *Management Science*, **22**, 1043-1047.

DEKKER, R., M.J. KLEIJN, AND A.G. DE KOK (1997). The break quantity rule's effect on inventory costs in a 1-warehouse, N-retailers distribution system. *International Journal of Production Economics*, to appear.

DIKS, E.B., AND M.C. VAN DER HEIJDEN (1997). Modeling stochastic lead times in multi-echelon systems. *Probability in the Engineering and Informational Sciences*, To appear.

DIKS, E.B., A.G. DE KOK, AND A.G. LAGODIMOS (1997). Multi-echelon systems: a service measure perspective. *European Journal of Operational Research*, **95**, 241-263.

DUNSMUIR, W.T.M., AND R.D. SNYDER (1989). Control of inventories with intermittent demand. *European Journal of Operational Research*, **40**, 16-21.

EIJS, M.J.G. VAN (1993). *Coordinated replenishment policies,* Ph.D. dissertation, Tilburg University, The Netherlands.

EPPEN, G., AND L. SCHRAGE (1981). Centralized ordering policies in a multi-warehouse system with random lead times and random demand. *Management Science*, **16**, 51-67.

FEARON, H.E., AND M. R. LEENDERS (1993). *Purchasing and materials management,* Irwin, Homewood Ill.

FEDERGRUEN, A. (1993). Centralized planning models, in Graves, S.C., A.H.G. Rinnooy Kan and P.H. Zipkin (eds.), *Logistics of production and inventory*, Handbooks in operations research and management science 4, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, Chapter 3, 133-173.

FEDERGRUEN, A.. AND P. ZIPKIN (1984). An efficient algorithm for computing optimal (s, S) inventory policies. *Operations Research*, **32**, 1268-1285.

FELLER, W. (1970). *An introduction to probability theory and ist applications*, John Wiley & Sons, New York.

FORRESTER, J.W. (1961). *Industrial Dynamics*, M.I.T. Press, Cambridge Massachusetts.

GALLIERS, R.D. (1992). *Information systems research: issues, methods and practical guidelines*, Oxford, England.

GALLIHER, H., P. MORSE, AND M. SIMOND (1959). Dynamics of two classes continuous-review inventory systems. *Operations Research*, **7**, 362-384.

GROSS, D., AND A. SORIANO (1969). The effect of reducing leadtime on inventory levels-simulation analysis. *Management Science*, **16**, B(61-76).

GUO, Y., AND R. GANESHAN (1995). Are more suppliers better? *Journal of Operational Research Society*, **46**, 892-895.

HADLEY, G., AND T.M. WHITIN (1963). *Analysis of inventory systems*, Prentice-Hall, Englewood Cliffs, NJ.

HALL, R.W. (1983). *Zero Inventories*, Dow Jones-Irwin, Homewood.

HARRIS, F.W. (1913). How many parts to make at once. *Factory, magazine of management*, **10**, 135-136.

HAYYA, J.C., D.P.C. CHRISTY AND A. PAN (1987). Reducing inventory uncertainty: reorderpoint system with two vendors. *Production and Inventory Management*, **28**, 43-49.

HEIJDEN, M.C. VAN DER (1993). *Performance analysis for reliability and inventory models*, Ph.D. dissertation, Free University of Amsterdam, The Netherlands.

HEUTS, R.M.J., AND J. DE KLEIN (1995). An (s,q) inventory model with stochastic and interrelated lead times. *Naval Research Logistics*, **42**, 839-859.

HILL, R.M. (1996). Order splitting in continuous review (Q, r) inventory models. *European Journal of Operational Research*, **95**, 53-61.

HONG, J.D., AND J.C. HAYYA (1992). Just-in-time purchasing: Single or multiple sourcing?. *International Journal of Production Economics*, **27**, 175-181.

HORDIJK, A., AND F.A. VAN DER DUYN SCHOUTEN (1986) On the optimality of (s, S)-policies in continuous review inventory models. *SIAM Journal on Applied Mathematics*, **46**, 912-929.

INDERFURTH K. (1996). *Modeling period review control for a stochastic product recovery problem with remanufacturing and procurement leadtimes*, Research memorandum, Faculty of Economics and Management, University of Magdeburg, Germany.

JANSSEN, F.B.S.L.P., A.G. DE KOK, F.A. VAN DER DUYN SCHOUTEN (1995). *Approximations for the delivery splitting model*, CentER discussion paper 9584, Tilburg University, The Netherlands.

JANSSEN, F.B.S.L.P., A.G. DE KOK (1996). *Demand management in a multi-stage distribution chain*, CentER Discussion paper 9639, Tilburg University, The Netherlands.

JANSSEN, F.B.S.L.P., R.M.J. HEUTS, AND A.G. DE KOK (1996). *The value of information in an (R,s,Q) inventory model*, CentER discussion paper 9621, Tilburg University, Tilburg, The Netherlands.

JANSSEN, F.B.S.L.P., A.G. DE KOK (1997a). *The optimal number of suppliers in an (s, Q) inventory system with order splitting*, CentER Discussion paper 9761, Tilburg University, The Netherlands.

JANSSEN, F.B.S.L.P., A.G. DE KOK (1997b). *The fill rate service measure in an (s, Q) inventory system with order splitting*, CentER Discussion paper 9762, Tilburg University, The Netherlands.

JANSSEN, F.B.S.L.P., A.G. DE KOK (1997). A two-supplier inventory model.*International Journal of Production Economics*, to appear.

JANSSEN, F.B.S.L.P., R.M.J. HEUTS, AND A.G. DE KOK (1998). On the (R, s, Q) inventory model when demand is modelled as a compound Bernoulli process. *European Journal of Operational Research*, **104**, 423-436.

JOHANSEN, S.G.J., AND A. THORSTENSON (1993). Optimal and approximate (Q, r) inventory policies with lost sales and gamma-distributed lead time. *International Journal of Production Economics*, **30**, 179-194.

JÖNSSON, H., AND E.A. SILVER (1987). Analysis of two-echelon inventory control system with complete redistribution. *Management Science*, **33**, 215-227.

KAPLAN, R.S. (1970). A dynamic inventory model with stochastic lead times. *Management Science*, **16**, 491-507.

KELLE, P., AND E.A. SILVER (1990a). Decreasing expected shortages through order splitting. *Eng. Costs and Production Economics*, **19**, 351-357.

KELLE, P., AND E.A. SILVER (1990b). Safety stock reduction by order splitting. *Naval Research Logistics*, **37**, 725-743.

KLEIJN, M.J., AND R. DEKKER (1997). *Using break quantities for tactical optimisation in multistage distribution systems.* Technical Report 9737/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands.

KLEMM, H., AND M. MIKUT (1972). *Lagerhaltungsmodelle, Theorie und Anwendungen,* Verlag Die Wirtschaft Berlin, Berlin.

KLEMM, H. (1974). Leiferbereitschaft und Vorratsnorm in Klemm, H and P Linke (eds.), *Lagerhaltungsmodelle, Neue Anwendungen, Erfahrungen und Erkenntnisse,* Verlag Die Wirtschaft Berlin, Berlin, Chapter 13, 250-269.

KOK, A.G. DE (1989). A moment-iteration method for approximating the waiting-time characteristics of the $GI/G/1$ queue. *Probability in the Engineering Informational Sciences*, **3**, 273-287.

Kok, A.G de (1991a). *Basics of inventory management: part 2 The (R,S)-model*, Research memorandum, FEW 521, Department of Economics, Tilburg University, Tilburg, The Netherlands.

Kok, A.G de (1991b). *Basics of inventory management: part 3 The (b,Q)-model*, Research memorandum, FEW 522, Department of Economics, Tilburg University, Tilburg, The Netherlands.

Kok, A.G. de (1991c). *Basics of inventory management: part 5 The (R,b,Q)-model*, Research memorandum, FEW 524, Department of Economics, Tilburg University, Tilburg, The Netherlands.

Kok, A.G. de (1993). *Back-order lead time behaviour in (s,Q) inventory models with compound renewal demand*, Research Report TUE/BDK/KBS/93-21, Eindhoven University of Technology, The Netherlands.

Kok, A.G. de (1996). *Analysis of (s,nQ) installation stock policies in divergent multi-echelon networks*, Research Report TUE/TM/LBS/96-09, Eindhoven University of Technology, The Netherlands.

Kok, A.G. de, D.F. Pyke, and M.P. Baganha (1996). *The undershoot of the reorder-level in an (s,nQ) model and its relation to the replenishment order size distribution*, Research Report TUE/TM/LBS/96-07, Eindhoven University of Technology, The Netherlands.

Kok, A.G. de, and K. Inderfurth (1997). Nervousness in inventory management: Comparison of basic control rules. *European Journal of Operational Research*, **103**, 55-82.

Lagodimos, A.G. (1992). Multi-echelon service models for inventory systems under different rationing policies. *International Journal of Production Research*, **30**, 939-958.

Larson, P.D. (1989). An inventory model which assumes the problem away: note on Pan and Liao. *Production and Inventory Management Journal*, **30** (4), 73-74.

Lau, H.S., and A.H. Lau (1994). Coordinating two suppliers with offsetting lead time and price performance. *Journal of Operations Management*, **11**, 327-337.

Lau, H.S., and L.G. Zhao (1993). Optimal ordering policies with two suppliers when lead times and demand are all stochastic. *European Journal of Operational Research*, **68**, 120-133.

Lee, H.L., and C. Billington (1992). Managing supply chain inventory: pitfalls and opportunities. *Sloan Management Review*, **33**, 65-73.

Lee, H.L., and C. Billington (1993). Material management in decentralized supply chains. *Operations Research*, **41**, 835-847.

LEE, H.L., AND S. NAHMIAS (1993). Single-product, single-location models, in Graves, S.C., A.H.G. Rinnooy Kan and P.H. Zipkin (eds.), *Logistics of production and inventory*, Handbooks in operations research and management science 4, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, Chapter 1, 3-55.

MARTIN, A.J. (1990). *Distribution Resource Planning*, Oliver Wight Ltd. Publications, 2-nd ed., Essex Junction.

MCLAUGHLIN, C.P., G. VASTAG, AND D.C. WHYBARK (1994). Statistical inventory control in theory and practice. International Journal of Production Economics, **35**, Iss 1-3, 161-169.

MOON, I., AND S. CHOI (1994). The distribution free continuous review inventory system with a service level constraint. *Computers and Industrial Engineering*, **27**, 209-212.

ORLICKEY, J. (1975). *Material Requirements Planning*, McGraw Hill, New York.

PAN, A.C., AND C. LIAO (1989). An inventory model under just-in-time purchasing agreements. *Production and Inventory Management Journal*, **30** (1), 49-52.

PRESS, W. H., B.P. FLANNERY, S.A. TEUKOLSKY AND W.T. VETTERLING (1992). *Numerical Recipes in C*, Cambridge, England.

RAMASESH, R.V. (1990). Recasting the traditional inventory model to implement just-in-time prurchasing. *Production and Inventory Management Journal*, **31**, 71-75.

RAMASESH, R.V. (1991). Procurement under uncertain supply lead times - a dual sourcing technique could save costs. *Engineering Costs and Production Economics*, **21**, 59-68.

RAMASESH, R.V., J.K. ORD, J.C. HAYYA AND A. PAN (1991). Sole versus dual sourcing in stochastic lead-time (s, Q) inventory models. *Management Science*, **37**, 428-443.

SAHIN, I. (1983). On the continuous review (s,S) inventory model under compound renewal demand and random lead times. *Journal of Applied Probability*, **20**, 213 - 219.

SAHIN, I. (1990). *Regenerative inventory systems*. Springer-Verlag, New York.

SCARF, H.E. (1960). The optimality of (s, S) policies in the dynamic inventory problem, in Arrow, K.A, S. Karlin, and P. Suppes (eds.), *Mathematical methods in social sciences*, Stanford University Press, Stanford, CA.

SCHNEIDER, H. (1981). Effect of service-levels on order-points or order-levels in inventory models. *International Journal of Production Research*, **19**, 615-631.

SCHNEIDER, H., AND J.L. RINGUEST (1990). Power approximation for computing (s,S) policies using service level. *Management Science*, **36**, 822-834.

SCULLI, D., AND S.Y. WU (1981). Stock control with two suppliers and normal lead times. *Journal of Operational Research Society*, **32**, 1003-1009.

SILVER, E.A. (1970). Some ideas related to inventory control of items having erratic demand patterns. *Journal of the Canadian Operations Research Society*, **8**, 87-100.

SILVER, E.A., AND R. PETERSON (1985). *Decision Systems for Inventory Management and Production Planning, 2nd edition*, Wiley, New York.

SILVER, E.A., D.F. PYKE AND R. PETERSON (1998). *Inventory management and production planning and scheduling, third edition*, Wiley, New York.

STAHL, M.J., AND D.W. GRIGSBY (1992). *Strategic management for decision making.* PWS-Kent Publishing Company, Boston.

TERSINE, R.J. (1994). *Priciples of inventory and materials management*, Prentice-Hall, 4-ed, New Jersey.

TIJMS, H.C., AND H. GROENEVELT (1984). 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.

TIJMS, H.C. (1994). *Stochastic Models: An Algorithmic Approach*, Wiley, Chichester.

VASTAG, G., J.D. KASARDA AND T. BOONE (1994). Logistical support for manufacturing agility in global markets. *International Journal of Operations & Production management*, **14**, 73-85.

VAUGHAN, T.S. (1995). The effect of sampling variability on statistical order point computations. *Production and Inventory Management Journal*, **36**, 84 - 87.

VEINOTT, A.F. (1966). On the optimality of (s, S) inventory policies: new conditions and a new proof. *SIAM Journal on Applied Mathematics*, **14**, 1067-1083.

WILLEMAIN, T.R., C.N. SMART, J.H SHOCKOR,AND P.A. DESAUTELS (1994). Forecasting intermittent demand in manufacturing : a comparative evaluation of Crostons' method. *International Journal of Forecasting*, **10**, 529-538.

WHITT, W. (1979). Approximating a point process by a renewal process: a general framework. Bell Laboratories.

WHITT, W. (1986). Approximating a point process by a renewal process I: two basic methods. *Operations Research*, **30**, 125-147.

ZHAO, L.G., AND H.S. LAU (1992). Reducing inventory costs and choosing suppliers with order splitting. *Journal of Operational Research Society*, **43**, 1003-1008.

ZIPKIN, P. (1986a). Stochastic leadtimes in continuous-time inventory models. *Naval Research Logistics Quarterly*, **33**, 763-774.

ZIPKIN, P. (1986b). Inventory service-level measures: convexity and approximations. *Management Science*, **32**, 975-981.

Samenvatting

Dit proefschrift is gewijd aan het management van één-produkt voorraadsystemen. Voorraadmanagement omvat het besturen en beheersen van de goederenstroom in en uit een voorraadpunt. Sinds de ontdekking van de beroemde 'EOQ'-formule worden in de praktijk wiskundige beschrijvingen van voorraadsystemen veelvuldig gebruikt om de effectiviteit en efficiëntie van voorraadsystemen te verbeteren. Bij de beschrijving van voorraadsystemen maken we in dit proefschrift onderscheid tussen: de bestel-strategie voor het aanvullen van voorraden; de beschrijving van de klanten en hun vraagomvang; en de levertijden van de aanvulorders.

In hoofdstuk 2 geven we een uitgebreide beschrijving van deze 3 componenten. Vervolgens definiëren we een aantal performancematen die de effectiviteit en efficiëntie van het voorraadsysteem beschrijven. De maten die gerelateerd zijn aan de effectiviteit worden aangeduid met (klant-)service-maten en de maten die gerelateerd zijn aan efficiëntie met (boete-)kosten-maten. Voor het bepalen van de optimale parameters van de bestelstrategie wordt in het algemeen een keuze gemaakt tussen twee criteria: het service-criterium en het kosten-criterium. Bij het service-criterium worden de totale bestel- en voorraadkosten geminimaliseerd met als nevenvoorwaarde dat een bepaalde klant-service moet worden gerealiseerd. Het kosten-criterium minimaliseert de som van de bestel-, voorraad-, en boetekosten. Omdat er aanzienlijke verwarring bestaat over de relatie tussen deze twee criteria wordt deze relatie voor het (s, Q)-bestelsysteem grondig onderzocht in paragraaf 2.3. Omdat de klantenvraag en levertijden meestal het beste gemodelleerd worden door stochastisch variabelen, behandelen we in paragraaf 2.4 een zeer veelzijdige en numeriek interessante klasse van verdelingsfuncties, namelijk de mengsels van Erlang verdelingen.

In hoofdstuk 3 analyseren we een (R, s, Q) voorraadsysteem waarbij de vraag van klanten gemodelleerd wordt door een samengesteld Bernoulli proces, ofwel met een bepaald vaste kans is de vraag gedurende een tijdseenheid positief en in het andere geval is de vraagomvang gelijk aan 0. Deze modelwijze is vooral interessant voor vraagprocessen met een onregelmatig karakter. Gedurende het laatste decennium heeft de informatie technologie zich sterk ontwikkeld, wat onder andere de registratie van vraagtransacties heeft verbeterd. Hierdoor komt zonder extra kosten informatie beschikbaar over vraag per dag in plaats van vraag per maand of per kwartaal. De kans dat er vraag is gedurende een maand is voor de meeste produkten 1, echter de kans dat er vraag is op een willekeurige dag is voor veel produkten significant lager dan 1. Er zijn twee belangrijke redenen om voor dit soort situaties een samengesteld Bernoulli proces te gebruiken. Ten eerste, indien men de verdelingsfunctie van de vraag per dag modelleert als een continue verdelingsfunctie op $[0, \infty)$, dan verliest men informatie over de puntmassa in nul. Ten tweede, indien onregelmatige vraagprocesen worden voorspelt, dan blijkt dat het decompositie idee van Croston beter voorspelt dan exponetiële vereffening, toegepast op de niet-gesepareerde data.

In hoofdstuk 4 bekijken we wederom het (R, s, Q) voorraadsysteem. In dit hoofdstuk vergelijken we een discrete-tijdsmodellering met een continue-tijdsmodellering. Discrete-tijdsmodellen nemen aan dat de tijd is opgedeeld in disjuncte tijdsintervallen met gelijke lengte, bijvoorbeeld een dag of een week. Verder wordt aangenomen dat de vraag in twee

opeenvolgende tijdsperiode onderling onafhankelijk is. Bij continue tijdsmodellen worden de tussenaankomsttijden van klanten en de vraagomvang van klanten apart gemodelleerd, en wordt er verondersteld dat deze onderling onafhankelijk zijn. De hamvraag van dit hoofdstuk is hoe goed een discrete-tijdsmodel fungeert als model voor een een continue-tijdsmodel. Voor het continue-tijdsmodel nemen we het samengesteld vernieuwingsproces. Het blijkt dat voor situaties dat de variatie coefficiënt van de tussenaankomsttijden van klanten lager is dan 1, de vraag in twee op eenvolgende tijdsperiode negatief gecorreleerd is, wat resulteert in een hoger service-niveau dan berekend. Voor variatie coefficiënten groter dan een is er positieve correlatie tussen de vraag in twee opeenvolgende perioden, wat leid tot een lagere service-niveau dan berekend.

In hoofdstukken 5 en 6 behandelen we bestelstrategiën die gebruik maken van meerdere leveranciers. In de literatuur worden deze strategiën aangeduid met 'order splitting' strategiën. 'Order splitting' kan worden gecombineerd met de bekende bestelstrategiën zoals de (s, Q)- en (s, S)-bestelstrategie. Bij 'order splitting' worden aanvulorders opgesplitst en simultaan bestelt bij meerdere leveranciers. De essentie van order splitting is de reductie in de leveronzekerheid. Indien een leverancier niet tijdig kan leveren, door stakingen of produktieproblemen, dan heeft dat bij order splitting geen dramatische gevolgen. Verder zal door onderlinge concurrentie tussen de leveranciers de inkoopprijs mogelijk dalen en de kwaliteit van de produkten mogelijk verbeteren. Hoofdstuk 5 behandelt een variant op 'order splitting'. Voor het bestellen van de aanvulorders worden twee leveranciers gebruikt. De inkoopprijs van produkten bij leverancier 1 zijn lager dan bij leverancier 2, echter de lengte van de levertijd van leverancier 1 is veel hoger dan van leverancier 2. Leverancier 1 zou bijvoorbeeld een producent kunnen zijn, en leverancier 2 een groothandel. Iedere periode wordt een vaste hoeveelheid bestelt bij de leverancier 1. Indien de economische voorraad na bestelling bij leverancier 1 lager is dan een bepaald aanvulniveau dan wordt een bestelling geplaatst bij leverancier 2 zodanig dat de economisch voorraad wordt aangevuld tot het aanvulniveau. Voor het bepalen van de optimale parameter van de bestelstrategie wordt het service-criterium gebruikt. Uit de numerieke resultaten blijkt dat alleen in gevallen van extreem onregelmatige vraagprocessen een groot gedeelte van het totale inkoopvolume bij leverancier 2 wordt bestelt om zo op korte-termijn veranderingen van de vraag te kunnen reageren.

In hoofdstuk 6 gaan we uitgebreid in op 'order splitting' in een (s, Q) voorraadsysteem. De vraag wordt gemodelleerd door een samengesteld vernieuwingsproces. De levertijden van de afzonderlijke leveranciers zijn stochastisch en hebben dezelfde verdelingsfunctie, die komt uit de klasse van mixed Erlang verdelingen. Verder wordt er geen restrictie gelegd op het aantal leveranciers. Door deze milde aannames worden al de tot nu toe behandelde modellen in de literatuur betreffende 'order splitting' speciale gevallen van het model behandelt in dit proefschrift. Veel van de modellen uit de literatuur gebruiken de kans op tekorten aan het einde van een bestelcyclus als service-maat bij het servicecriterium. Echter de opsplitsing van aanvulorders verandert de structuur van de bestelcyclus aanzienlijk. Daarom bespreken we deze service-maat in paragraaf 6.3, waarvoor blijkbaar een herdefinitie noodzakelijk is. In de literatuur wordt bijna geen aandacht geschonken aan twee in de praktijk erg veel gebruikt service maten, namelijk: de fractie van de vraag direct geleverd uit voorraad; en de kans dat een klant niet uit voorraad belevert kan worden. Deze worden in paragraaf 6.4 uitgebreid behandeld. Tenslotte onderzoeken we wat het optimaal aantal leveranciers is aan de hand van een boetekosten model. Voor het optimaliseringsprobleem spelen de bestelkosten, die afhangen van het aantal leveranciers, een belangrijke rol. De algemeenheid van het model en de snelheid van het algoritme maken het mogelijk om voor de meeste praktijksituaties snel oplossingen te genereren. Het management kan op deze manier snel en adequaat onderzoeken of 'order splitting' kosten efficiënt is.

In hoofstuk 5 en 6 hebben we onderzocht hoe met behulp van meerdere leveranciers de goederenstroom van leverancier naar het voorraadpunt kan worden verheersd. In hoofdstuk 7 en 8 wordt onderzocht hoe de goederenstroom uit het voorraadpunt naar klanten kan worden beheersd. Hiervoor worden twee strategiën behandeld, namelijk, 'large order overflow' en 'delivery splitting'. Hoofdstuk 7 behandelt 'large order overflow'. 'Large order overflow' is bedoeld om de goederenstromen in een multi-echelon voorraadsysteem te stroomlijnen. Indien een klant arriveert bij een voorraadpunt in het multi-echelon systeem die een vraag heeft die groter is dan een bepaalde drempelwaarde, dan wordt deze klant doorgestuurd naar alternatief voorraadpunt. Het alternatieve voorraadpunt wordt zo gekozen dat de vraag van de doorgestuurde klanten voor dat voorraadpunt kleine of normale klanten zijn. Door 'large order overflow' toe te passen wordt de variatie van de vraagomvang van klanten gereduceerd, en dus kan de veiligheidsvoorraad van het betreffende voorraadpunt worden verlaagd om dezelfde service te realiseren. Ten eerste behandelen we een serieel netwerk met 2 (s, Q) voorraadsystemen. Bij het tweede depot wordt 'large order overflow' toegepast. Voor dit model is een algoritme ontwikkeld voor het bepalen van de optimale parameters van de (s, Q)-strategie en de drempelwaarde bij het tweede voorraadpunt. De extra transportkosten van de klanten die worden doorgestuurd naar het eerste voorraadpunt spelen bij dit optimaliseringsprobleem een belangrijke rol. Verder worden indicaties gegeven van de optimale waarde van de drempelwaarde bij diverse waarden voor de inputparameters. Indien de optimale drempelwaarde naar oneindig gaat is 'large order overflow' niet kosten-efficiënt. Het ander extreem is de situatie dat de optimale drempelwaarde naar 0 gaat, in dit geval worden alle klanten doorgestuurd naar het alternatieve voorraadpunt en wordt het betreffende voorraadpunt overbodig. Tenslotte behandelen we een divergent 2-echelon systeem. Voor dit model kan een gelijksoortig algoritme, als voor het serieel netwerk afgeleid is, gebruikt worden. Hiermee geven we aan dat we ook in staat zijn om 'large order overflow' door te rekenen in een wat meer complex multi-echelon netwerk.

In hoofdstuk 8 behandelen we de ander uitleverbeheersingsstrategie: 'delivery splitting'. 'Delivery splitting' kan worden toegepast op een afzonderlijk voorraadpunt. Bij delivery splitting word een klant met een vraagomvang groter dan een zeker drempelwaarde niet direct beleverd, maar wordt zijn vraag gesplitst in deel-leveringen die met vaste tussenpozen worden geleverd. Op deze manier wordt, net als bij 'large order overflow', de variantie van de uitlevergrootte gereduceerd. Deze strategie heeft een sterke overeenkomst met 'order splitting'. Indien bij een voorraadpunt 'delivery splitting' wordt toegepast, dan komen de gevraagde (aanvul)-orders van klanten met een vraag groter dan de drempelwaarde gespreid in de tijd binnen bij die klant. Dit kan deze klant dan ervaren als een impliciete vorm van 'order splitting'. Het belangrijke verschil tussen deze twee strategiën is dat bij 'order splitting' het proces wordt bekeken vanuit het oogpunt van het depot richting de leverancier, en bij 'delivery splitting' vanuit het voorraadpunt richting de klanten. Voor een compleet beeld van de voor- en nadelen van 'delivery splitting' moeten we de impact van deze strategie op de klanten meenemen. Hiervoor kunnen we dan een aangepaste vorm van 'order splitting' gebruiken. Aan het einde van hoofdstuk 8 wordt hiervan een voorbeeld behandeld. Verder wordt in dit hoofdstuk ingegaan op de impact van delivery splitting op het voorraadpunt waar 'delivery splitting' wordt toegepast. Omdat 'delivery splitting' resulteerd in gecorreleerde uitleveringen ontwikkelen we twee methoden voor het bepalen van de performance maten, waarbij de eerste methode de correlatie verwaarloost maar numeriek eenvoudig is. Tenslotte merken we op dat door delivery splitting informatie beschikbaar komt over uitleveringen in de toekomst. In paragraaf 8.5 behandelen we een model waarin deze informatie wordt gebruikt bij het maken van aanvulbeslissingen. Het kern-idee is om niet op de economische voorraad te sturen maar op het 'available-to-promise' voorraad niveau, dat gedefinieerd is als de economische voorraad minus de bekende uitlevering gedurende de levertijd. Voor dit model worden voor diverse inputparameter de additionele besparingen doorgerekend ten opzichte van het model met 'delivery splitting' waarbij geen rekening wordt gehouden met informatie over bekende toekomstige uitleveringen. Het blijkt dat dit nog behoorlijke extra besparingen kan opleveren.

Center for Economic Research, Tilburg University, The Netherlands
Dissertation Series

No.	Author	Title
1	P.J.J. Herings	Static and Dynamic Aspects of General Disequilibrium Theory; ISBN 90 5668 001 3
2*	Erwin van der Krabben	Urban Dynamics: A Real Estate Perspective - An institutional analysis of the production of the built environment; ISBN 90 5170 390 2
3	Arjan Lejour	Integrating or Desintegrating Welfare States? - a qualitative study to the consequences of economic integration on social insurance; ISBN 90 5668 003 x
4	Bas J.M. Werker	Statistical Methods in Financial Econometrics; ISBN 90 5668 002 1
5	Rudy Douven	Policy Coordination and Convergence in the EU; ISBN 90 5668 004 8
6	Arie J.T.M. Weeren	Coordination in Hierarchical Control; ISBN 90 5668 006 4
7	Herbert Hamers	Sequencing and Delivery Situations: a Game Theoretic Approach; ISBN 90 5668 005 6
8	Annemarie ter Veer	Strategic Decision Making in Politics; ISBN 90 5668 007 2
9	Zaifu Yang	Simplicial Fixed Point Algorithms and Applications; ISBN 90 5668 008 0
10	William Verkooijen	Neural Networks in Economic Modelling - An Empirical Study; ISBN 90 5668 010 2
11	Henny Romijn	Acquisition of Technological Capability in Small Firms in Developing Countries; ISBN 90 5668 009 9
12	W.B. van den Hout	The Power-Series Algorithm - A Numerical Approach to Markov Processes; ISBN 90 5668 011 0
13	Paul W.J. de Bijl	Essays in Industrial Organization and Management Strategy; ISBN 90 5668 012 9

* Copies can be ordered from ThelaThesis,Prinseneiland 305, 1013 LP Amsterdam, The Netherlands, phone + 31 20 6255429; fax: +31 20 6203395; e-mail: office@thelathesis.nl

No.	Author	Title
14	Martijn van de Ven	Intergenerational Redistribution in Representative Democracies; ISBN 90 5668 013 7
15	Eline van der Heijden	Altruism, Fairness and Public Pensions: An Investigation of Survey and Experimental Data; ISBN 90 5668 014 5
16	H.M. Webers	Competition in Spatial Location Models; ISBN 90 5668 015 3
17	Jan Bouckaert	Essays in Competition with Product Differentiation and Bargaining in Markets; ISBN 90 5668 016 1
18	Zafar Iqbal	Three-Gap Analysis of Structural Adjustment in Pakistan; ISBN 90 5668 017 x
19	Jimmy Miller	A Treatise on Labour: A Matching-Model Analysis of Labour-Market Programmes; ISBN 90 5668 018 8
20	Edwin van Dam	Graphs with Few Eigenvalues - An interplay between combinatorics and algebra; ISBN 90 5668 019 6
21	Henk Oosterhout	Takeover Barriers: the good, the bad, and the ugly; ISBN 90 5668 020 x
22	Jan Lemmen	Financial Integration in the European Union: Measurement and Determination; ISBN 90 5668 021 8
23	Chris van Raalte	Market Formation and Market Selection; ISBN 90 5668 022 6
24	Bas van Aarle	Essays on Monetary and Fiscal Policy Interaction: Applications to EMU and Eastern Europe; ISBN 90 5668 023 4
25	Francis Y. Kumah	Common Stochastic Trends and Policy Shocks in the Open Economy:Empirical Essays in International Finance and Monetary Policy; ISBN 90 5668 024 2
26	Erik Canton	Economic Growth and Business Cycles; ISBN 90 5668 025 0
27	Jeroen Hoppenbrouwers	Conceptual Modeling and the Lexicon; ISBN 90 5668 027 7

No.	Author	Title
28	Paul Smit	Numerical Analysis of Eigenvalue Algorithms Based on Subspace Iterations; ISBN 90 5668 026 9
29	Uri Gneezy	Essays in Behavioral Economics; ISBN 90 5668 028 5
30	Erwin Charlier	Limited Dependent Variable Models for Panel Data; ISBN 90 5668 029 3
31	Rob Euwals	Empirical Studies on Individual Labour Market Behaviour; ISBN 90 5668 030 7
32	Anurag N. Banerjee	The Sensitivity of Estimates, Inferences, and Forecasts of Linear Models; ISBN 90 5668 031 5
33	Frans A. de Roon	Essays on Testing for Spanning and on Modeling Futures Risk Premia; ISBN 90 5668 032 3
34	Xiangzhu Han	Product Differentiation, Collusion and Standardization; ISBN 90 5668 033 1
35	Marcel Das	On Income Expectations and Other Subjective Data: A Micro-Econometric Analysis; ISBN 90 5668 034 X
36	Jeroen Suijs	Cooperative Decision Making in a Stochastic Environment; ISBN 90 5668 035 8
37	Talitha Feenstra	Environmental Policy Instruments and International Rivalry: A Dynamic Analysis; ISBN 90 5668 036 6
38	Jan Bouwens	The Use of Management Accounting Systems in Functionally Differentiated Organizations; ISBN 90 5668 037 4
39	Stefan Hochguertel	Households' Portfolio Choices; ISBN 90 5668 038 2
40	Henk van Houtum	The Development of Cross-Border Economic Relations; ISBN 90 5668 039 0
41	Jorg Jansen	Service and Inventory Models subject to a Delay-Limit; ISBN 90 5668 040 4
42	F.B.S.L.P. Janssen	Inventory Management Systems: control and information issues; ISBN 90 5668 041 2

FRED JANSSEN grad

University of Technology. Between 1994 and 1998 he was a Ph.D. student at the Cooperation Center Tilburg and Eindhoven Universities (SOBU). He worked at the department of Operations Planning and Control in Eindhoven and at the department Econometrics in Tilburg. His main research interests lie in the field of inventory control.

This dissertation addresses the management of inventory systems. The thesis starts with an exposition on mathematical models that can be used in inventory theory. Then we deal with some information issues related to the demand process. Namely, how to control products that have intermittent demand. Moreover, we investigated the impact of data collection on the customer performance. Next, we investigated to what extend multiple-sourcing can lead to improvements of the inventory system. Finally two demand management strategies are investigated for smoothing demand. The first re-routes large customer orders to alternative stockpoint, whereas the second strategy splits a customer order in a time-phased delivery scheme.

ISBN 90 5668 041 2

CentER Dissertation