Exact Methods for Multi-echelon Inventory Control

Incorporating Shipment Decisions and Detailed Demand Information

Stenius, Olof

2016

Citation for published version (APA):
Stenius, O. (2016). *Exact Methods for Multi-echelon Inventory Control: Incorporating Shipment Decisions and Detailed Demand Information*.

Total number of authors:
1

Exact Methods for Multi-echelon Inventory Control

– Incorporating Shipment Decisions and Detailed Demand Information

Olof Stenius

LUND
UNIVERSITY

Organization	Document name: Doctoral thesis
LUND UNIVERSITY Department of Industrial Management & Logistics Division of Production Management	**Author**: Olof Stenius
	Date of dissertation: 13 May, 2016

Title and subtitle: Exact Methods for Multi-echelon Inventory Control – Incorporating Shipment Decisions and Detailed Demand Information

Abstract:

Recent advances in information technologies and an increased environmental awareness have altered the prerequisites for successful logistics. For companies operating on a global market, inventory control of distribution systems is often an essential part of their logistics planning. In this context, the research objective of this thesis is:

To develop exact methods for stochastic inventory control of multi-echelon distribution systems incorporating shipment decisions and/or detailed demand information.

The thesis consists of five scientific papers (Paper I, II, III, IV and V) preceded by a summarizing introduction. All papers study systems with a central warehouse supplying a number of non-identical local warehouses (retailers) facing stochastic demand. For given replenishment policies, the papers provide exact expressions for evaluating the expected long-run system behavior (e.g., distributions of backorders, inventory levels, shipment sizes and expected costs) and present optimization procedures for the control variables.

Paper I and II consider systems where shipments from the central warehouse are consolidated to groups of retailers and dispatched periodically. By doing so, economies of scale for the transports can be reached, reducing both transportation costs and emissions. Paper I assumes Poisson customer demand and considers volume-dependent transportation costs and emissions. The model involves the possibility to reserve intermodal (train) capacity in combination with truck transports available on demand. For this system, the expected inventory costs, the expected transportation costs and the expected transport emissions are determined. Joint optimization procedures for the shipment intervals, the capacity reservation quantities, the reorder points and order-up-to levels in the system are provided, with or without emission considerations. Paper II analyses the expected costs of the same system for compound Poisson demand (where customer demand sizes may vary), but with only one transportation mode and fixed transportation costs per shipment. It also shows how to handle fill rate constraints.

Paper III studies a system where all stock points use installation stock (R,Q) ordering policies (batch ordering). This implies that situations can occur when only part of a requested retailer order is available at the central warehouse. In these situations, the models in existing literature predominantly assume that available units are shipped immediately (partial delivery). An alternative is to wait until the entire order is available before dispatching (complete delivery). The paper introduces a cost for splitting the order and evaluates a system where optimal choices between partial and complete deliveries are made for all orders. In a numerical study it is shown that significant savings can be made by using this policy compared to systems which exclusively use either partial or complete deliveries.

Paper IV shows how companies can benefit from detailed information about their customer demand. In a continuous review base stock system, the customer demand is modeled with independent compound renewal processes at the retailers. This means that the customer inter-arrival times may follow any continuous distribution and the demand sizes may follow any discrete distribution. A numerical study shows that this model can achieve substantial savings compared to models using the common assumption of exponential customer inter-arrival times.

Paper V is a short technical note that extends the scope of analysis for several existing stochastic multi-echelon inventory models. These models analyze the expected costs without first determining the inventory level distribution. By showing how these distributions can be obtained from the expected cost functions, this note facilitates the analysis of several service measures, including the ready rate and the fill rate.

Key words: Inventory, Multi-echelon, Stochastic, Shipment decisions, Detailed demand information		
Language: English	**Number of pages**: 214	**ISRN**: LUTMDN/TMIO-1013-SE
ISBN print: 978-91-7623-794-6	**ISBN pdf**: 978-91-7623-795-3	

Signature _____ Date 2016-04-06

Exact Methods for Multi-echelon Inventory Control

– Incorporating Shipment Decisions
and Detailed Demand Information

Olof Stenius

Exact Methods for Multi-echelon Inventory Control – Incorporating Shipment Decisions and Detailed Demand Information

To Nasti, Edda and Lykke

Abstract

Recent advances in information technologies and an increased environmental awareness have altered the prerequisites for successful logistics. For companies operating on a global market, inventory control of distribution systems is often an essential part of their logistics planning. In this context, the research objective of this thesis is:

> *To develop exact methods for stochastic inventory control of multi-echelon distribution systems incorporating shipment decisions and/or detailed demand information.*

The thesis consists of five scientific papers (Paper I, II, III, IV and V) preceded by a summarizing introduction. All papers study systems with a central warehouse supplying a number of non-identical local warehouses (retailers) facing stochastic demand. For given replenishment policies, the papers provide exact expressions for evaluating the expected long-run system behavior (e.g., distributions of backorders, inventory levels, shipment sizes and expected costs) and present optimization procedures for the control variables.

Paper I and II consider systems where shipments from the central warehouse are consolidated to groups of retailers and dispatched periodically. By doing so, economies of scale for the transports can be reached, reducing both transportation costs and emissions. Paper I assumes Poisson customer demand and considers volume-dependent transportation costs and emissions. The model involves the possibility to reserve intermodal (train) capacity in combination with truck transports available on demand. For this system, the expected inventory costs, the expected transportation costs and the expected transport emissions are determined. Joint optimization procedures for the shipment intervals, the capacity reservation quantities, the reorder points and order-up-to levels in the system are provided, with or without emission considerations. Paper II analyses the expected costs of the same system for compound Poisson demand (where customer demand sizes may vary), but with only one transportation mode and fixed transportation costs per shipment. It also shows how to handle fill rate constraints.

Paper III studies a system where all stock points use installation stock (R,Q) ordering policies (batch ordering). This implies that situations can occur when only part of a requested retailer order is available at the central warehouse. In these situations, the models in existing literature predominantly assume that

available units are shipped immediately (partial delivery). An alternative is to wait until the entire order is available before dispatching (complete delivery). The paper introduces a cost for splitting the order and evaluates a system where optimal choices between partial and complete deliveries are made for all orders. In a numerical study it is shown that significant savings can be made by using this policy compared to systems which exclusively use either partial or complete deliveries.

Paper IV shows how companies can benefit from detailed information about their customer demand. In a continuous review base stock system, the customer demand is modeled with independent compound renewal processes at the retailers. This means that the customer inter-arrival times may follow any continuous distribution and the demand sizes may follow any discrete distribution. A numerical study shows that this model can achieve substantial savings compared to models using the common assumption of exponential customer inter-arrival times.

Paper V is a short technical note that extends the scope of analysis for several existing stochastic multi-echelon inventory models. These models analyze the expected costs without first determining the inventory level distribution. By showing how these distributions can be obtained from the expected cost functions, this note facilitates the analysis of several service measures, including the ready rate and the fill rate.

Keywords: Inventory, Multi-echelon, Stochastic, Shipment decisions, Detailed demand information

Acknowledgements

The completion of this thesis would not have been possible without the help and support of a number of people. First and foremost, I wish to thank my supervisor, prof. Johan Marklund, co-author of Paper I and II in this thesis. He has helped me at all stages of the research process. He has guided me towards interesting research paths and kept my plans and expectations realistic. His comments and feedback on my drafts have developed me as a writer and improved the end product greatly. He has also generously offered me a place to stay during my visits in Lund. I have been very lucky to have such a supportive and generous supervisor.

Second, I wish to thank my former colleague and close friend, Christian Howard, co-author of Paper III. His role, especially during the first years of the PhD process, was absolutely critical. His support helped me through the roughest parts of my PhD process. Our long discussions have guided me in finding interesting research directions and helped me manage my work. Christian and his family have also generously taken care of me during my stays in Lund.

I would like to express the deepest appreciation to my assistant supervisor, professor Sven Axsäter, co-author of paper I. He is an inspiring role model and has given me valuable advice and feedback throughout the process. My assistant supervisors, associate prof. Peter Berling and assistant prof. Fredrik Olsson, have generously offered me comments and advice, improving my research. The successful collaboration with Gönül Karaarslan and prof. Ton de Kok, which resulted in paper II, turned out to be a turning point in my PhD process. Thank you also Fredrik Eng Larsson, Lina Johansson, Birgitta Olvegart and all other current and former colleagues at the department.

I would also like to express my gratitude to prof. Markku Kuula and prof. Ari Vepsäläinen and the department of Information and Service Economy at Aalto University, Finland. You have generously provided me with a place to work and a friendly work environment during my stay in Finland. Prof. Søren Glud

Contents

1. List of Papers

- **Paper I – Sustainable Multi-echelon Inventory Control with Shipment Consolidation and Volume Dependent Freight Costs**
  Stenius, O., J. Marklund, S. Axsäter
  Department of Industrial Management and Logistics, Lund University
  (2016) Submitted

- **Paper II –Exact Analysis of Divergent Inventory Systems with Time-Based Shipment Consolidation and Compound Poisson Demand**
  Stenius, O., G. Karaarslan, J. Marklund, A. G. de Kok
  Department of Industrial Management and Logistics, Lund University
  (2016) *Operations Research*, forthcoming

- **Paper III – Partial or Complete Deliveries in Two-echelon Inventory Systems?**
  Howard, C., O. Stenius
  Department of Industrial Management and Logistics, Lund University
  (2016) Under revision

- **Paper IV – Divergent Two-echelon Inventory System with Compound Renewal Demand**
  Stenius, O.
  Department of Industrial Management and Logistics, Lund University
  (2016) Submitted

- **Paper V – A Note on Solution Procedures for a Class of Two-echelon Inventory Problems**
  Stenius, O.
  Department of Industrial Management and Logistics, Lund University
  (2016) Submitted

2. Introduction

Globalization, advances in information technologies and increased environmental awareness have transformed the prerequisites for logistics planning during the last decades. With geographically larger systems to control and an intensified focus on emissions, the need to control large distribution systems incorporating shipment aspects has increased. Developments in information technologies have also facilitated this control in many ways. It is now in the Big Data era cheaper and easier to share information within the supply chain and many companies also collect vast amount of detailed information about their businesses, including their customer demand patterns.

In this thesis, these issues are studied from an inventory control perspective. Inventory control is a crucial function in many companies. Product availability is a prerequisite for sales and shortages can decrease companies' revenues directly by loss of revenues and increasing administrative costs, as well as indirectly by damaging the companies' reputation. On the other hand, there are considerable costs associated with keeping materials and products in stock; cost of tied up capital, storage costs, costs obsolesce, etc. Creating sound rules and methods to balance these aspects is key to successful inventory management. For distribution systems, inventory control becomes challenging especially when future demand is uncertain. This stochasticity is present in most businesses, but it tends to be accentuated in distribution systems with expensive low demand items, such as spare parts. This thesis presents four inventory control models for distribution systems in four different research papers focusing on incorporating shipment decisions (papers I, II and III) and/or on incorporating more detailed demand information (paper II and IV). Paper V is a technical note that provides some general relationships between the costs and the distributions of the amount of inventory kept at a stock point for a group of existing inventory control models.

This introductory chapter describes and summarizes the background, the aim, the research performed and the main contributions of Papers I-V. One of the purposes of this chapter is to make the research in the papers comprehensible for a somewhat broader audience. Section 2.1 defines the research objectives. Section 2.2 discusses the research methodology used to reach the results. In order for the reader to understand the context of the performed research, Section 2.3 presents a conceptual overview of the field of inventory control and Section 2.4 summarizes the most relevant existing contributions related to the performed research. Section 2.5 describes the modelling features, Section 2.6 summarizes the contents of

Papers I-V and Section 2.7 highlights the contributions of the thesis. Finally Section 2.8 discusses future research. This introductory chapter is partly based on Stenius (2014), which is the licentiate thesis of the same author and includes earlier versions of Paper I, Paper II and Paper III.

2.1 Research Objectives

The research objective can be stated as follows:

To develop exact methods for stochastic inventory control of multi-echelon distribution systems incorporating shipment decisions and/or detailed demand information.

To clarify, let us define the terminology and concepts used in this objective statement. "*To develop... methods for stochastic inventory control*" means that, for given assumptions regarding the system (including the stochastic demand and the replenishment policies) methods are developed to evaluate expected costs and/or service levels and to optimize defined control variables (e.g. reorder points or shipment intervals). The concept "*exact*" is defined so that for given assumptions, all mathematical expressions and relationships are obtained without introducing any simplifying approximations. The "*distribution systems*" considered in this work consist of a central warehouse supplying a group of local warehouses (referred to as retailers). Descriptions of how the models in the papers are "*incorporating shipment decisions and/or detailed demand information*" are provided below.

Shipment decisions are included in the thesis in two different ways. Paper I and II investigate how a distribution system can be controlled using a time-based shipment consolidation policy. In these models, the central warehouse consolidates shipments to groups of retailers and dispatches them periodically. The models determine how often shipments should be made, and how much stock should be kept at each location under different cost structures. Paper I considers the option to reserve capacity on a cheaper or more environmentally friendly transportation mode. Apart from the costs, Paper I also evaluates the emissions during transportations. These are considered in the cost optimization, for instance by adding a constraint on the maximum expected transport emissions allowed.

Paper III, also includes shipment decisions, but has a different focus. In a distribution system with batch ordering, it is investigated when partial or complete deliveries from the central warehouse should be used. This decision is explained by the following example: A retailer orders five units from the central warehouse. At this time, the warehouse only has three units available and a

replenishment is arriving in two days. The warehouse thus needs to make a decision between either shipping the three available units immediately (partial delivery), or waiting until all five units ordered are available for shipment (complete delivery). By introducing a cost for splitting the order, this delivery decision is optimized. The total system costs are then evaluated for the system where this decision is made for every retailer order, and these costs are compared to the costs when only partial or only complete deliveries are used.

Paper IV incorporates detailed demand information. More specifically, it focuses on systems that possess detailed information about the time between customer arrivals (the inter-arrival times) and customer demand sizes. The inter-arrival times are allowed to follow any continuous distribution and the demand sizes follow any discrete distribution. Especially for the inter-arrival times, this constitutes a difference to the majority of existing literature on exact analysis of distribution inventory systems, which typically assume exponential inter-arrival times. The exponential distribution has many analytical advantages, but this paper shows that approximating the inter-arrival time distribution with exponential times can be costly.

Paper V does not present a new model. Instead, it shows how to obtain additional information from a group of existing inventory control models. More specifically, for a specific set of systems (analyzed in several previous research articles), this note shows how the inventory level distribution (i.e. the stock on hand minus the backorders) can be obtained from the cost function. Through this information, several performance measures can be determined. For instance, for many systems, it provides information about the proportion of demand that can be satisfied directly from stock (i.e. the fill rate).

2.2 Research Methodology

As stated in Section 2.1 above, the research objective is to develop exact methods for stochastic inventory control methods. For this purpose, mathematical models that characterize and solve real life problems are developed. Stochastic inventory control modeling is typically achieved by applying and expanding theories from the fields of probability theory, optimization, queuing theory, control theory, statistics, computer science and programming on problems formulated based on knowledge in logistics, economics and business administration. This is also the case for this thesis.

The modeling process is typically divided into three steps, which are repeated iteratively (see, for example, Hillier and Lieberman, 2010, and Axsäter and Marklund, 2010).

In the first step, the structure of the system is determined and necessary assumptions made. The model can either aim at being general, describing a problem found in various companies, or be case-specific, i.e. adjusted to a certain environment. The models in this thesis are all general, aiming at being applicable in many distribution systems. Note however, that when applied, these general models usually need to be adapted to fit the company's specific needs. In this first step, it is important to determine the level of detail, and decide on which features of reality that need to be captured in the model. If the model is too detailed and complex, it becomes very difficult to obtain any useful results. However, if the model is over-simplified, important characteristics of the problem are excluded and the results are misleading.

In the second step a solution to the problem is generated. This is done by applying and developing tools from the mathematical fields listed above. For the problems in this thesis, the main challenges for the analysis stem from the stochasticity of the demand. This uncertainty makes the relationships between different stochastic variables, such as customer demand at certain retailers and the amount of stock kept at a warehouse, rather complex. Numerous tools from the field of probability theory are used to analyze these relationships. In order to optimize the control parameters, methods from optimization theory are applied.

In the third step, the results are validated. In other words, it is assured that the presented model and results describe the original problem accurately. Validation can be separated into external and internal validation. The external validation concerns the extent to which the model makes a good representation of real world problems. In this thesis, the external validation is primarily based on validated models in existing literature. The relevance of the proposed models is assured by extending these inventory models in well-motivated, relevant directions, e.g. based on discussions with industry partners. The internal validation evaluates the correctness of the generated results of the model. Because all models in this thesis are analyzed exactly, the internal validation is primarily considered in the proofs of the mathematical expressions in the papers. Apart from the mathematical solutions in the papers, the models also consist of a computer program that solves the problems according to the presented solution. In order to verify that the programming is performed correctly, the analytical results are compared with simulated results for sufficiently many different kinds of problems. If the generated solution is incorrect, one has to correct the program, or in some cases, return to step 2 to correct the solution procedure or even to step 1 and adjust the models.

When stated that the presented models are analyzed exactly, this means that after the system structure is determined and the assumptions describing the system behavior are set in step 1, no approximations are introduced to facilitate

the cost evaluation and optimization of the model in step 2. Exact models will obviously generate the best solution for the specified problem, but it is not an obvious choice for complex Operations Research problems. For large complex systems with e.g. many stock points, many different products and high demand rates, exact evaluation and optimization techniques can become too computationally cumbersome to implement. For these systems, fast and good heuristic (approximate) solutions can be preferable. The motivation for focusing on exact solutions, apart from the fact that they create stable implementable solutions for smaller systems, is that good heuristic solutions often are based on exact solutions and that the exact solution can serve as a point of reference for heuristic solutions. Exact solutions can also provide valuable insights on the dynamics of the system performance and optimality of different policies.

Apart from exact and heuristic solutions, another possible way to evaluate inventory systems is by simulation. Simulations have the advantage of being flexible in the sense that it is possible to model complex systems and policies. There are, however, also drawbacks with this methodology. Firstly, optimization via simulation search is often very time-consuming. The simulation run times tend to be quite long to get good solutions. Also, one simulation run will typically provide the result for one set of parameters only and the parameter sets grow rapidly as the systems become larger and more complex. As a result, finding the optimal policy through simulation is often not practical. Secondly, it is more difficult to achieve understanding of the dynamics in the system, when the mathematical relationships are left unexplored. Simulation is used as a tool also in this thesis, but only for the verification of the results.

2.3 Overview of Basic Inventory Control Concepts and Theory

In order to facilitate the reading of the papers in the thesis, this section introduces some of the basic concepts used in the field of inventory control theory. As the models in this thesis are based on stochastic demand (future demand is uncertain), this section is restricted to these types of models and problems. For a more thorough understanding of the broad field of inventory control/management we refer to, for example, Silver et al. (1998), Zipkin (2000) or Axsäter (2006).

The field of Inventory control deals with managing material flows in companies and supply chains and is traditionally focused on the questions; when should new material be ordered, produced or shipped? And how much material should be ordered/produced/shipped? One of the most commonly known inventory control problems is the *newsvendor* problem (Edgeworth, 1888). This problem studies a newsvendor, who wishes to optimize the number of newspapers

he should procure at the beginning of a day. The daily demand is uncertain and he wishes to sell as many papers as possible, while avoiding lots of excess papers unsold at the end of the day. This relatively simple inventory control problem considers a single period, and only deals with the question of how much material to order. However, when allowing for material to be stored and sold later, the decision space grows and includes the question of when to order as well. In the remainder of this section we will describe features that define different inventory control problems, with a focus on features characterizing the problems studied in this thesis.

2.3.1 Structure

The structure, or the topology, of the problem describes the stock points included in the system, and how they are connected. Namely, from where each stock point receives its replenishments and how material flows through the system. The simplest and most commonly studied structure is the single-echelon inventory system consisting of only a single stock point. A single-echelon system is illustrated in Figure 1, where the arrows illustrate the material flow and the triangle illustrates a stock point. This stock point receives demand from customers, satisfies this demand if possible, and replenishes stock from an outside supplier (or an internal production unit).

Figure 1. Single-echelon system

Multi-echelon systems feature multiple connected stock points. The simplest multi-echelon system is the serial system. Here every stock point has only one immediate predecessor and one immediate successor. These types of systems can be seen in many production facilities, where the connections between the stock points can be seen as production processes. A three-echelon serial system is illustrated in Figure 2.

Figure 2. Three-echelon serial system

The systems studied in this thesis are divergent distribution inventory systems. Here, each stock point has only one predecessor, but can have many successors. As the name indicates, distribution systems are common in companies handling physical distribution of products. Often they have central warehouses

located in conjunction to their production units, and local warehouses closer to the different markets. Conceptually, distribution systems can also be found in production units where a raw material or component is diversified to several products. In this case, the central warehouse corresponds to the stock of raw materials and the local warehouses to the stocks of each product. An example of a two-echelon distribution system is illustrated in Figure 3.
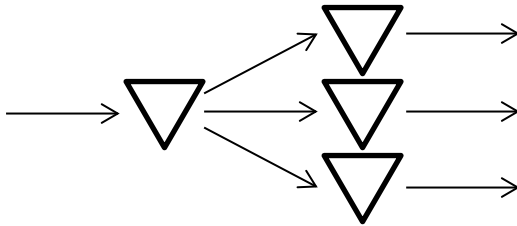
Figure 3. Example of a two-echelon distribution system

Other multi-echelon systems include assembly systems, where different components are assembled to an end product. In assembly systems each stock point only has one successor but may have many predecessors. There also exist other, more general, multi-echelon structures, where stock points may have many predecessors as well as many successors. These systems exist for instance in production facilities where many different products are assembled from partly different components or in companies that both assembles and distributes products. Some systems allow for material flow between stock points at the same level. For instance, a retailer facing stock outs may receive units from another retailer nearby. This feature, referred to as lateral transshipments, is excluded from this thesis, but might be an interesting direction for future research. For an overview of models handling lateral transshipments, see, for example, Paterson et al. (2011).

2.3.2 Lead Times

The replenishment lead times are usually defined as the time it takes from a replenishment order is placed until the products are available at the ordering inventory location. A large part of the lead time often consists of transportation or production time, or a combination of the two. However, it also includes the time for order placement, picking, loading and receiving activities. Sometimes there exist restrictions on when transportation or production can take place, for instance periodic transportation/production schemes. The replenishment lead time then also includes the waiting time for capacity to become available. In multi-echelon systems the lead time also includes the time spent waiting for products or components to become available at stock points upstream. Sometimes the term

transportation lead time is used to define the time for order placement, picking, loading, transportation and receiving activities, thus excluding the time the system waits for units becoming available or the time until scheduled transportations leave.

Often, in practice, the lead time varies, but if the variation is small, it is common from an inventory modelling perspective to assume that the lead time is constant. This facilitates the analysis of more complex problems. The transportation lead time variation is also excluded from the models studied in this thesis. However, they do consider the waiting time for transportation capacity to become available and variation caused by and stock-outs at the preceding stock points, which in many multi-echelon systems cause a major part of the lead time variation.

2.3.3 Customer Demand and Review Periods

The representation of the customer demand differs depending on the type of problem investigated. In the single-period problem illustrated by the newsvendor problem, each day/period is controlled separately. In these systems the customer demand is characterized by the probability distribution of the demand in one period. This distribution is usually assumed to be known. Apart from the single-period problems, the inventory control problems may be classified as periodic review problems or continuous review problems.

Periodic review problems indicate that the stock levels are examined periodically (e.g. once every day) and that replenishment orders only can be placed when a review is performed. The demand in periodic review problems is usually characterized by the distribution of the demand in one period, often with the assumption that the demands in different periods are independent.

The inventory problems considered in this thesis are so called continuous review problems. This means that demand is observed the moment it is received and necessary actions can be taken immediately. Even though the demand is monitored continuously in all models in this thesis, the problems in Paper I and Paper II are related to the periodic review models as shipments from the central warehouse to the retailers are performed periodically.

In continuous review systems the demand is often characterized by a stochastic process, which means that the probabilities of when customers will arrive and how much they will order are known (in some simpler systems, the distribution of the demand during a replenishment lead time is sufficient to analyze the system). The most commonly used process is the Poisson process. When customer demand follows a Poisson process, the time between consecutive customer arrivals (the inter-arrival times) are independent and exponentially distributed and each customer orders one unit. An important property of the

Poisson process is that the time until the next customer arrival is independent of the time since the previous customer arrived. Partially because of this independency, the Poisson process has several analytical advantages. The Poisson process is characterized by a single parameter; the arrival intensity (i.e. the expected amount of customers arriving per time unit). Moreover, the variance of the demand during a given period is always equal to the mean (the variance-to-mean ratio is one). This may not always be a good representation of the actual demand, seen in the system. However, in many systems dealing with, for instance, spare parts, the Poisson process is known to describe the real demand process of some products quite well. The demand is assumed to follow a Poisson process in paper I and III of this thesis.

For a more flexible demand representation, the compound Poisson process can be used. Here, the customers still arrive independently with exponential inter-arrival times. However, each customer can demand any number of units (independently of the quantity of the other customers). The compound Poisson distribution is characterized by the arrival intensity of the customers and the distribution of the amount of units demanded by an arbitrary customer (the demand sizes). The compound Poisson process can handle any variance-to-mean ratio of the demand per period larger than or equal to one. In Paper II of the thesis, the demand is assumed to follow a compound Poisson process.

An even more general demand structure is considered in Paper IV of the thesis. Here, the times between customer arrivals may follow any continuous distribution and the demand sizes follow any discrete distribution. The inter-arrival times and demand sizes are still assumed to be independent of each other. This process corresponds to a compound renewal processes. This is a very flexible demand structure, and it can handle any variance-to-mean ratio of the demand per period. However, when the inter-arrival times are non-exponential, some of the analytical advantages of the Poisson process are lost. For instance, the demands during two consecutive time periods are no longer independent, which causes challenges when analyzing complex multi-echelon systems.

2.3.4 Performance Measures

The performance measures assess the quality of a specific system setting, and thereby set the goal of the inventory control. In the example with the newsvendor it is commonly assumed that there is a fixed purchasing price and a fixed selling price per newspaper copy. The performance measure is thus the expected profit of the newsvendor. This type of measure is not unusual for inventory control problems, the objective is often either to maximize the profit or to minimize the costs (if the revenue is given). Common costs that are included in these

optimization problems are holding costs, ordering or setup costs, and backorder or shortage costs.

The holding cost includes all the costs for storing the product (e.g. opportunity cost for tied-up capital, material handling, storage, damage and obsolescence, insurance, and taxes). It is often assumed that the expected holding cost is proportional to the expected inventory on hand.

The ordering cost (or setup cost) is a fixed cost associated with placing a replenishment order. These fixed costs can occur, for instance, in production, during transportation or in administration of the order. The existence of ordering costs is one of the main reasons why production and transportation is performed in batches, but there can also be practical reasons for batching (for instance packaging sizes).

The backorder costs (or penalty/shortage costs) occur when the company is unable to fulfill customer demand. Consequences related to unsatisfied customer demands vary significantly between different companies and industries. There are situations when the customer will go to a competitor if the product is not available (lost sales), and others where the customer is willing to wait until the product becomes available (complete backordering). The models in Papers I-V all assume complete backordering. When a unit is backordered it incurs backorder costs for the company. These costs include administrative costs, possible price reductions offered to the customer, and the trade mark and brand damage caused by the shortage. Often these costs are increasing, the longer the customer has to wait. It is therefore common to assume that the expected backorder costs are proportional to the expected number of backorders.

The backorder cost per unit and time unit may be difficult to quantify in practice. Many companies therefore control their inventories using service levels instead. Common service levels include the ready rate and the fill rate. The ready rate is defined as proportion of time with positive stock on hand, while the fill rate is the proportion of demand that can be satisfied directly from stock. If service level constraints are used, the inventory control objective is typically to minimize the (holding and ordering) costs while assuring that the service constraints are met. Many performance measures, such as the ready rate and the fill rate, can be determined by first analyzing the inventory level distribution. The inventory level is commonly defined as the amount of stock on hand minus the amount of backordered units. Paper V focus on this aspect by generalizing a group of models that previously only analyzed the costs, to also obtaining the inventory level distribution. In the thesis, Papers I, III and IV are based on backorder costs, while Paper II allows for either backorder costs or fill rate constraints.

2.3.5 Replenishment Policies

The replenishment policy specifies the rules according to which the replenishments are made. Sometimes the optimal replenishment policy is the result of the defined problem. For instance, in the newsvendor problem it can be shown that the optimal order quantity is $q^* = F^{-1}((p–c)/p)$ papers each day. Here F^{-1} is the inverse cumulative distribution of the demand per day, p is the selling price and c is the purchase price (in this variation of the problem unsold products have no salvage value). In more complex systems the optimal ordering policy is often unknown and one has to settle for a predetermined replenishment policy (hopefully) performing close to optimal.

In continuous time systems, the simplest replenishment policy is the (S–1,S) policy, also referred to as the base stock policy, one-for-one replenishments (in systems facing unit demand) or the order-up-to S policy (predominantly in periodic review systems). For this policy, every customer demand immediately triggers an order of the same size (in the continuous review case). This replenishment policy is also optimal in many systems where the demand in consecutive time periods are independent and there are no ordering or setup costs.

For single-echelon systems with fixed ordering costs, it has been shown that the (s,S) policy is optimal under very general conditions, see Iglehart (1963), Veinott (1966), Porteus (1971) and Zheng (1991). The (s,S) policy implies that as soon as the inventory position (the stock on hand + outstanding orders – backorders) drops to or below the order point, s, an order is placed to bring the inventory position up to the order-up-to level, S.

In some systems, there exist practical reasons (sizes of packages and load carriers) for replenishing in fixed batches. These systems typically use (R,Q) ordering policies. This implies that an order of Q units is placed as soon as the inventory position drops to, or below, the reorder point, R. Sometimes, several (n) simultaneous orders of Q units may be required to bring the inventory position above R again. In these cases, the (R,Q) policy is often denoted (R,nQ). In this section, however, we will use the notation (R,Q) in both cases. Note that, in continuous review models where customers order one unit at a time, the (s,S) policy is equivalent to an (R,Q) policy (with R=s and R+Q=S). Also, if s = S–1 in the (s,S) policy or Q = 1 in the (R,Q) policy, both policies become equivalent to the base stock policy.

For multi-echelon systems facing stochastic demand, the optimal replenishment policy is known only for a restricted set of problems. In serial systems where there are ordering costs only at the most upstream facility, it has been shown that it is optimal to order with an (s,S) policy at this facility and with (S–1,S) policies at all other facilities, see Clark and Scarf (1960) and Federgruen

and Zipkin (1984). There also exist some optimality results for assembly systems (see, Rosling 1989), but for most multi-echelon systems facing stochastic demand, the optimal replenishment policy is unknown. The majority of the existing literature on multi-echelon inventory control is thus focused on determining the costs (exactly or by heuristics) and optimizing system parameters under reasonable replenishment policies. This is also the case in this thesis.

The terms replenishment policy and ordering policy are often used synonymously. However, in the first three papers of this thesis, the replenishment policy consists of both an ordering policy from the downstream facility and a delivery policy of the upstream facility. As the administrative costs of placing orders become smaller (due to the development of information systems), the incentives for batch ordering (i.e. batching information about demand that has decreased) decrease. However, the incentives for batching physical products into consolidated deliveries (or production activities) still exist and are accentuated by increasing transportation costs and environmental concerns. It might therefore be beneficial to let the downstream locations share all their demand information with the upstream location and let the upstream location handle the consolidation (or batching) decisions. The upstream location has more information and can for instance consolidate shipments of different units and products to different downstream locations together in an efficient way (as studied in Paper I and II of the thesis).

Another aspect affecting the replenishments in multi-echelon distribution systems is the allocation policy at upstream locations. When several downstream locations have requested units from the same upstream stock point and there is not enough inventory available to satisfy all requests, the upstream stock point is faced with an allocation decision. A simple allocation rule commonly used in practice is the First-Come-First-Served (FCFS) policy. Here, the downstream facility that order first receives the first replenishment. This allocation policy is popular among practitioners as it is rather easy to implement and it seems "fair". The FCFS policy also has many analytical advantages and is the dominant allocation policy used in the literature on exact analysis of continuous review inventory systems. This is also the allocation policy chosen for the papers in this thesis. Howard and Marklund (2011) and Howard (2013) have investigated the benefits of using more sophisticated allocation policies in multi-echelon distribution systems with time-based shipment consolidation. The conclusion is that some savings can be attained, but in most cases the FCFS policy performs very well.

2.4 Literature Review

In order to understand the context in which the contributions of this thesis are made, this chapter presents a short overview of the most relevant existing multi-echelon inventory models. In line with the research in Papers I-V, we focus on exact analysis of stochastic models of distribution systems under continuous review. All of the papers discussed below study one-warehouse-multiple-retailer systems, but some of the models can be extended to several echelons.

One complicating matter in the analysis of these systems usually lies in the fact that the performance of the retailers is dependent on the central warehouse. More specifically, the stock outs (backorders) at the central warehouse cause delays in the replenishment lead time to the retailers. Analyzing this delay or the distribution of the backorders at the central warehouse is the key in most of the models below.

Simon (1971) analyzes a continuous review spare parts distribution system where the retailers face Poisson demand and all stock points applies (S-1,S) ordering policies. By analyzing the distribution of the amount of backorders at the central warehouse destined to a specific retailer at an arbitrary point in time, he is able to determine the distribution of the inventory level (the stock on hand – the backorders) at this retailer, and thereby the costs. Kruse (1979) extends this result to more than two echelons and Graves (1985) presents a framework for exact and approximate solutions based on this system.

Axsäter (1990) presents an alternative way to analyze the system in Simon (1971). His approach is however very different. He follows an arbitrary unit as it travels through the system and analyzes the expected holding and backorder costs that this unit incurs. He also presents a fast recursive procedure for evaluating the costs and optimizing the order-up-to levels. The drawback of his procedure is that it does not determine the inventory level distribution and thereby limits the range of alternative performance measures that may be considered. Paper V provides the inventory level distribution from the cost function of this system. The results of this note are also generalized to several other models that build on Axsäter (1990).

The base model in Axsäter (1990) is generalized to handle compound Poisson demand (customers can order more than one unit at the time) in Forsberg (1995). Paper IV analyzes a similar system, but for the case of compound renewal demand (when the customer inter-arrival times can follow a general distribution). Note however, that the analysis of Paper IV is based on a different methodology than Forsberg (1995), and also provides the inventory level distribution.

Systems where the stock points apply (R,Q) policies have also been widely studied. In these papers, two different central warehouse (R,Q)-policies have been considered; installation stock, and echelon stock policies. Installation stock

policies imply that all stock points place orders based on their own inventory position (stock on hand + outstanding orders – backorders). For stock points applying echelon stock policies, orders are placed based on the sum of the inventory position at your own stock point and all downstream stock points. For two-echelon systems this only differentiates the replenishment policy at the central warehouse.

Axsäter (1993a) studies an installation stock (R,Q) policy system where the retailers face Poisson demand. The exact solution is restricted to the case where the retailers are identical. This result is generalized to non-identical retailers in Forsberg (1997a) and Axsäter (1998). Forsberg (1997b) extends the result to the case where customer inter-arrival times are Erlang-distributed. Axsäter (2000) uses a different method for analysis, where he also determines the inventory level distributions, for a system where customer demand is compound Poisson distributed.

For echelon stock (R,Q) policies, systems facing compound Poisson demand are analyzed in Axsäter (1997). In a parallel work, Chen and Zheng (1997) use a method of analysis resembling the one in Simon (1971) and Graves (1985), and also analyze the inventory level distribution. This analysis is exact only in the case of Poisson demand. All of the (R,Q) systems cited above assume partial deliveries (available units are shipped as soon as possible from the central warehouse). Andersson (1999) relaxes this assumption by introducing a minimum delivery quantity from the central warehouse in an installation stock (R,Q) system. This is related to Paper III, that also analyzes delivery policies in installation stock (R,Q) systems.

During the last decades, the focus of exact analysis in distribution systems has shifted to more advanced replenishment policies. Marklund (2002) considers an alternative way of using the inventory information at the retailers. He introduces an (α_0, Q_0) ordering policy at the central warehouse, with the aim of synchronizing replenishments with future retailer orders. Also Moinzadeh (2002) uses the inventory information from the retailers when triggering orders at the central warehouse. His analysis is exact in the case of identical retailers and identical order quantities at all stock points. For this system he lets the central warehouse trigger orders based on when the inventory position at the retailers reaches a value s, which can be different from the retailer reorder point R. Axsäter and Marklund (2008) provide an optimal "position based" ordering policy, which means that they allow for any ordering policy at the central warehouse that is dependent on the inventory positions at all stock points in the system. Compared to the other systems cited in this section, this is the only analysis that relaxes the FCFS allocation assumption. However, also in this system, the allocation of a specific unit to a future demand is determined the moment this unit is ordered to

the central warehouse from the outside supplier (or production unit). The ordering policy in Axsäter and Marklund (2008) has a performance guarantee over all of the other policies cited above. A limitation of this policy is that the minimum order quantity at the central warehouse cannot be set larger than the smallest order quantity at any of the retailers.

Marklund (2006) examines a system where the retailers get advance information about future demands. More specifically, each customer demand entails a due date, when this specific demand should be satisfied. All stock points use base stock policies, but different allocation strategies are used in order to utilize the advance information.

More recently, Marklund (2011) studies consolidation of orders in a distribution system. The central warehouse use an (R,Q) ordering policy, while the retailers (facing Poisson demand) use (S-1,S) ordering policies to immediately transfer the demand to the central warehouse. The central warehouse consolidates the retailer orders across different products but also across different retailers within the same region in order to get economies of scale for the transports. This setting is also considered in Papers I and II in this thesis. In a related study, Gürbüz et al. (2007) considers a system where the central warehouse is in charge of ordering, allocations and distribution to the retailers, but it cannot keep stock itself, and acts as a cross-docking facility.

Of the papers studying periodic replenishments in distribution systems, Axsäter (1993b), Forsberg (1995), Graves (1996) and Shang et al. (2015) are more closely related to our work as they use FCFS allocations. For an overview of the literature on inventory control in distribution systems with periodic review, see, for example, Axsäter (2003) and Marklund and Rosling (2012).

2.5 Modelling Features of Papers I-V

This section summarizes the modelling characteristics of Papers I-V. It thereby positions the papers in relation to existing literature presented in Section 2.4 based on the model assumptions. As mentioned earlier, the models presented in this thesis extend the existing literature of exact stochastic divergent inventory models primarily in two directions; namely, including delivery decisions (Papers I, II and III) and incorporating detailed demand information Paper II and IV). Paper V is a technical note that extends the scope of analysis for several existing models focused only on cost analysis by showing how to obtain the inventory level distributions.

All of the models analyzed in this thesis have the same structure (a single central warehouse and multiple non-identical retailers). They all assume continuous review, FCFS allocations, linear holding costs and constant

transportation lead times. Table 1 summarizes other distinguishing modelling characteristics considered in this thesis. Recall that Poisson demand is a special case of compound Poisson demand, which in turn is a special case of the compound renewal demand. Similarly, the (S-1,S) policy is a special case of the (R,Q) policy.

Table 1. Modelling charasteristics of papers I-V.

Paper	I	II	III	IV	V
Demand structure					
Poisson	X		X		X
Compound Poisson		X			X
Compound renewal				X	X
Ordering policies					
(S-1,S) at all locations				X	X
(R,Q) at central warehouse and (S-1,S) at retailers	X	X			X
(R,Q) at all locations			X		X
Shipment policy from central warehouse					
Time-based consolidation	X	X			X
Alternative transportation modes	X				
Transport emissions considered	X				
Optimizing partial/complete deliveries			X		
Service measures					
Shortage costs	X	X	X	X	X
Fill rates		X			X

2.6 Summary of Papers I-V

This section summarizes the most relevant aspects of Papers I, II, III, IV and V separately. For each paper the motivation of the study, the description of the problem, the key features of the analysis and the most relevant results and conclusions are provided. The main contributions are highlighted in Section 2.7.

2.6.1   Paper I – Sustainable Multi-echelon Inventory Control with Shipment Consolidation and Volume Dependent Freight Costs

Fluctuating fuel prices and environmental concern has led to an increased interest in railway transportations of goods. This paper is the result of discussions with several companies that have central warehouses in central Europe and aspirations

to ship goods by train to local warehouses in, for instance, Scandinavia. More specifically, they are interested in implementing intermodal truck-train-truck solutions (referred to as the primary transportation option) to groups of retailers within the same geographical area. Current market conditions for intermodal truck-train-truck transports typically dictate a solution where fixed capacity is reserved on trains with a preset shipment interval. This paper studies this issue by including realistic transportation costs and emissions in an inventory control model of a distribution system where shipments are consolidated periodically.

The considered model consists of a central warehouse supplying a number of retailers that each faces independent Poisson demand. The warehouse uses an (R,Q) policy to replenish from an outside supplier. The retailers use (S-1,S) ordering policies to immediately transfer the demand information to the warehouse. This is motivated by the diminishing costs of placing orders, as discussed earlier.

For economies of scale and to reduce the environmental impact, retailer orders are consolidated at the central warehouse and shipped to groups of retailers periodically. This means that available units ordered by retailers in a specific retailer group (retailer group k) are dispatched every T_k time units (where T_k denotes the shipment interval to this retailer group). For at least some of the retailer groups there is an opportunity to reserve capacity on a primary intermodal transportation option. When dispatched, this reserved capacity (w_k) is used to maximum extent and excess units are shipped with an alternative transportation option (truck) directly to the retailers. Figure 4 illustrates an example of the structure of the model where there are three retailers belonging to two retailer groups.
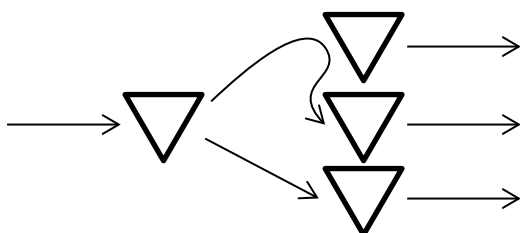
Figure 4. Example of the generic system considered in Paper I and Paper II.

Initially, a solution is presented for the single item case, where the transportation times of the two transportation options are equal. The analysis is then extended both to multi-item systems and systems where the transportation times are different for the two shipment options that may be used.

The model considers linear holding costs at each stock point (per unit and time unit) and linear backorder costs at the retailers. There are fixed costs

associated with each shipment leaving, depending on how much capacity that is reserved. There are also fixed costs for each truck in use (even if only part of this truck is used) and costs per unit for transporting with both transportation options. The transport emissions in the system are modeled by fixed emissions dependent on the capacity reservation, as well as emissions per unit transported and per load carrier in use, analogously to the cost parameters.

If the transportation times of the two shipment options are equal, the recursive method for exact evaluation of the expected inventory holding and backorder costs in Marklund (2011) is applicable. Marklund (2011) assumes a fixed cost for each shipment leaving, regardless of the amount of units shipped. In our work (presented in Paper I), we consider more general transportation costs and emissions, and allow for different transportation times. This is done by obtaining the distribution of the amount of units on each shipment. With this distribution known, it is straightforward to extend the model to handle any transportation structure where the transportation costs and emissions are dependent on the shipment size.

The paper presents a method to exactly evaluate the expected costs and transport emissions for this system. Based on this analysis, it is shown how to jointly optimize the shipment intervals to each retailer group (the T_k values), the amount of capacity to reserve on the intermodal train transports (the w_k values), and the reorder points and order-up-to levels in the system (the R and the S values). The optimization is explained for three different scenarios; (i) the emissions are not considered, (ii) there is a fixed cost per unit of emission, and (iii) there is a constraint on the maximum expected amount of emission incurred in the system, θ.

In a numerical example based on realistic cost and emission figures it is illustrated how the expected total cost $TC^*(\theta)$ depends on the emission target θ when optimizing scenario (iii) The example illustrates that relatively large emission reductions can be achieved with only marginal cost increases. However, larger emission decreases become expensive (see Figure 5).
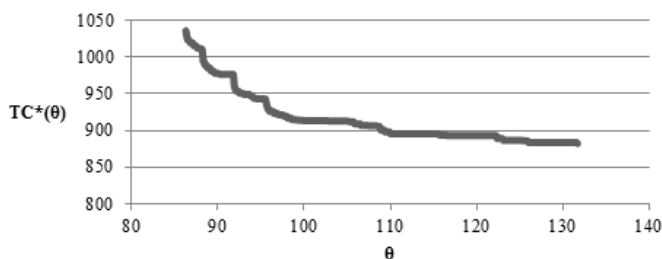
Figure 5. Expected total cost for varying θ values.

2.6.2 Paper II – Exact Analysis of Divergent Inventory Systems with Time-based Shipment Consolidation and Compound Poisson Demand

This paper investigates a system similar to the one considered in Paper I and Marklund (2011). There is a central warehouse supplying N retailers and shipments are consolidated and dispatched to groups of retailers periodically (see Figure 4, above). The main distinction from Paper I and Marklund (2011) is that the demand is now assumed to follow a compound Poisson process (note that the Poisson process assumed in Paper I and Marklund, 2011, is a special case of the compound Poisson process). This demand process can handle different variance to mean ratios and is thereby applicable in a much broader array of real problems. This paper also generalizes the model to handle fill rate constraints (i.e. constraints on the proportion of demand satisfied immediately from stock on hand) apart from backorder costs.

Another distinction compared to Paper I is that the shipment costs now only consist of fixed costs for each scheduled shipment (independent of volume) and costs per unit for each transported unit. Also, it is assumed that there is only one transportation option per retailer group and thus no reservation of capacity or difference in transportation times. These assumptions are realistic in many cases where the transports are bought from an outside (third or fourth party) logistics provider. This transportation cost structure simplifies the analysis of the shipment costs significantly. The main analytical challenges therefore lie in the evaluation of the long run expected inventory levels, which will provide the systems holding costs, backorder costs and fill rates.

As mentioned earlier, the complicating matter in the analysis of inventory levels in distribution systems is that the performance of the central warehouse affects the retailers. A shortage at the central warehouse can delay the shipment to the retailer and the replenishment lead time for a retailer order thus depends on whether there are units available at the central warehouse or not. In this paper this dependency is considered by obtaining the exact distribution of the amount of backorders at the central warehouse destined to a specific retailer i, B_i (when a shipment is leaving to this retailer). This distribution has previously been determined for the Poisson case in e.g. Simon (1971). However, the customer demand sizes in the compound Poisson case add additional complexity. The analysis of the backorders at the central warehouse is performed using a new approach, which can be used to solve also other types of inventory control problems in distribution systems (for instance, versions of this methodology is used in Papers III and IV).

It is a well-known result that the inventory level at an arbitrary point in time t_0 (= the stock on hand – the backorders) is equal to the inventory position (= inventory level + outstanding orders) at time $t_0 - L_0$ (a replenishment lead time,

L_0, earlier) minus the demand in time interval $(t_0 - L_0, t_0]$. The novelty in the proposed approach lies in tracking the *nominal inventory position*, which is a new concept introduced in the paper. The *nominal inventory position* is defined as the inventory position at $t_0 - L_0$ minus all demand at the central warehouse after $t_0 - L_0$. Thereby, the *nominal inventory position* (when positive) will serve as a measure of how many units the central warehouse still can satisfy before time t_0. A possible sample path of the *nominal inventory position* is illustrated in Figure 6. When the inventory level is negative (equal to $-x$) at time t_0, we know that there are exactly x backorders. Because of the FCFS assumption we also know that it will be the last x units ordered from the central warehouse before time t_0 that will be backordered at time t_0. By analyzing when the *nominal inventory position* is brought to a negative value (when the first backordered units at time t_0 are ordered by a retailer), the distribution of the backorders at the central warehouse destined to a specific retailer, B_i, can be analyzed. This constitutes the backbone of the analysis in this paper.
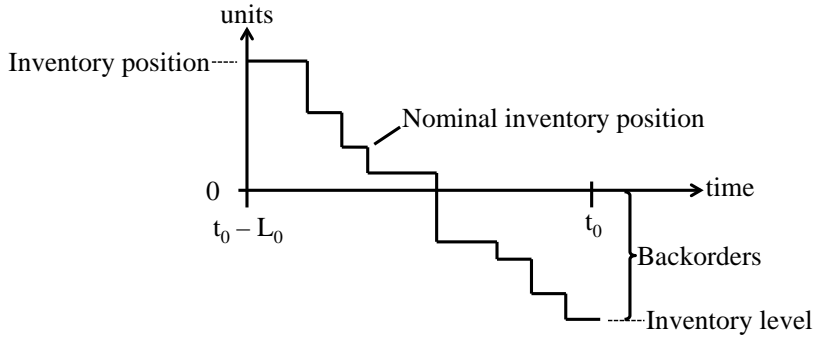
Figure 6. A sample path of the nominal inventory position.

With the distribution of backorders destined to retailer i, B_i, known, it is possible to determine the distribution of the inventory levels at this retailer. By conditioning on B_i at the moment a shipment is leaving to this retailer, the inventory level distribution a transportation lead time later at retailer i can be determined. Analogously, the inventory level during the following replenishment interval can be determined conditioning on B_i. After obtaining the distribution of the inventory levels, it is straightforward to analyze the expected costs and fill rates in the system. Based on these results, an optimization procedure is presented where the control parameters (the shipment intervals, the reorder points and the order-up-to levels) are jointly optimized (the order quantities at the central warehouse is assumed given). The costs are proven to be convex in the retailer order-up-to levels for a given warehouse reorder point and fixed shipment intervals. The optimization procedure is thus based on finding lower and upper optimality bounds on the warehouse reorder point and the shipment intervals. The

optimal solution is found by searching within these bounds and using the convexity of the retailer order-up-to levels. The proposed analytical methods for cost evaluation and the optimization procedure are applicable in both single- and multi-item systems.

2.6.3 Paper III – Partial or Complete Deliveries in Two-echelon Inventory Systems?

This paper also focuses on the shipment strategies and the shipment cost structure at the central warehouse, but for a different system and from another perspective. Here, an inventory distribution system consisting of a central warehouse and a number of retailers facing Poisson demand, where all stock points use (R,Q) policies (an example of the structure can be seen in Figure 3, above) is examined. The (R,Q) policy (fixed batch ordering) is one of the most commonly used replenishment policies in practice and it has also been widely studied in the literature, see, for example, Axsäter (2000). In distribution systems with batch ordering, situations can occur where only part of a retailer order is available at the central warehouse. In these situations, the warehouse can choose to dispatch the available units immediately and dispatch the remaining units as soon as they arrive to the central warehouse. This is referred to as a *partial delivery*. If the majority of the ordering cost is connected with the placement of an order, this can be a reasonable choice. If, however, there are substantial costs associated with splitting the delivery of an order, it can be reasonable to wait until the entire order is available and ship all units at ones, i.e. a *complete delivery*.

This paper introduces a cost, θ_i, of splitting an order to retailer i, and analyzes and compares three different delivery policies; a Partial Delivery policy (PD policy), a Complete Delivery policies (CD policy), and a Mixed State-Dependent policy (MSD policy). In the PD policy, only partial deliveries are used, and in the CD policy only complete deliveries are used. In the MSD policy, a cost optimization between a partial or a complete delivery is performed for each delivery. The cost optimization is based on information about how many units that are backordered and when the ordered units will become available for shipment. Note that, regardless of the delivery policies, the central warehouse allocates the units according to a FCFS policy.

The existing literature predominantly assumes PD policies, see Axsäter, (1993a, 1997, 1998, 2000), Chen and Zheng (1997) Forsberg (1997a,b) and Marklund (2002). In this current paper it is however proven that the MSD policy has a performance guarantee over both the PD and the CD policies (i.e. the costs for the MSD policy are at least as low as for the other policies). Also, it is shown that when the costs of splitting orders are sufficiently large, no partial deliveries will be made and the CD policy becomes equivalent to the MSD policy. Thus, the

CD policy has a performance guarantee over the PD policy in these situations. Note, however, that even if the cost of splitting orders is zero ($\theta_i = 0$), situations can occur, when it is beneficial to use complete deliveries.

For all three policies, the expected costs in the system are evaluated exactly. Because the MSD policy contains a state-dependent decision (dependent on when outstanding orders will arrive to the warehouse), the cost analysis for this policy is more complex than for the stationary PD and CD policies. The analysis utilizes a modified version of the *nominal inventory position* introduced in paper II. By further exploring the properties of the *nominal inventory position* it is possible to keep track of when orders have been placed (and will arrive to the warehouse). This enables the evaluation of the probabilities for different delivery decisions to occur. As a result, the distributions of the inventory levels and, consequently, the expected costs of the system can be obtained. With the cost analysis in place, optimization procedures for the reorder points (the R values) for the three policies are attained.

In a small numerical study, the costs and control parameters for the different policies are evaluated (see Paper III for details of the study). The study consists of 32 problems, which are all optimized for all three policies. For these problems, the expected costs for the PD policy are on average 5.8 % higher than under the MSD policy and the costs for the CD policy is 5.9 % higher than for the MSD policy. The maximum cost increase of using the PD policy instead of the MSD policy is 26.6% and the corresponding value for the CD policy is 17.9 %. This implies that the costs for choosing the wrong delivery policy can be significant. The study also indicates that when the costs for splitting orders increases, there is a tendency to keep more inventory at the central warehouse (under the MSD policy). This can be explained by the fact that the handling cost penalizes situations when there is not enough stock at the central warehouse. This implies that under our more general cost structure, there should be more stock kept at the central warehouse than what the existing multi-echelon literature suggests (see e.g. Axsäter 2003). The majority of the inventory will, however, still be kept at the retailers.

2.6.4 Paper IV – Divergent Two-echelon Inventory Systems with Compound Renewal Demand

The fourth paper of the thesis focuses on detailed demand information in distribution systems. Many companies today have access to exhaustive information about their business. For logistics planning this means that they possess detailed information about the demand processes of their customers. In order to benefit from this information, models are required that handle more detailed demand structures.

As stated earlier, most existing models of distribution systems assume that the inter-arrival times follow exponential distributions, resulting in Poisson or compound Poisson demand processes. This can be a good representation of reality if the customers arrive independently of each other. In other systems, for example for some critical spare parts, it is not unusual to have a local stock point in conjunction with every large customer. In these cases, increasing or decreasing failure rates of the spare parts triggering orders or customers' batching of orders can make the inter-arrival times far from exponential.

In the presented model, the customer inter-arrival times can follow any continuous distribution and the demand sizes of the customers can follow any discrete distribution. By assuming that the inter-arrival times and demand sizes are independent also across the retailers, the retailer demand processes constitute independent compound renewal processes. The replenishments in the model are made according to continuous review (S-1,S) policies at all stock points and all stock points apply FCFS allocation, complete backordering and partial deliveries. There are holding costs per unit and time unit at all stock points and backorder costs per unit and time unit at all retailers.

For this model, the exact inventory level distributions at all stock points in steady state are obtained. This is done by examining how the inventory levels at a given time depend on the customer demands during the previous replenishment lead times. The *nominal inventory position* from paper II is used to divide the analysis into different cases. By obtaining the distributions of the demand at the retailers during one or two consecutive time period(s) at a retailer, the inventory level distributions and the costs can be determined. From these distributions the expected total costs are evaluated and a recursive procedure to optimize the order-up-to levels (the S-values) at all stock points is presented.

In a numerical study the optimal system behavior is examined when the customer inter-arrival times are either gamma or Weibull distributed. Loosely speaking, the gamma distribution can be used to approximate customer arrivals, when the customers batch their orders, while the Weibull distribution is commonly used in reliability theory to model increasing or decreasing failure rates of spare parts (see, for instance, O'Connor and Kleyner, 2012). For the 120 problem scenarios tested, the expected cost increase of approximating these distributions with exponential distributions is evaluated. Moreover, as the gamma distribution has some analytical advantages over the Weibull distribution, it is examined whether the gamma distribution can be used to approximate the Weibull distributions.

The numerical study focuses on cases where the coefficient of variation (the standard deviation divided by the mean) of the inter-arrival times, ρ, is smaller than or equal to 1. Note that, for the exponential distribution, ρ is always

equal to 1. The focus on smaller ρ values is motivated by the fact that increasing failure rates or customers' batching of orders generates ρ values less than 1. Also, ρ is required to be smaller than 1 in order to model variance-to-mean ratios of the demand per time unit smaller than 1. This is relevant as the commonly used compound Poisson process only can model variance-to-mean ratios of the demand per time unit greater than or equal to 1. The details of the numerical study are found in the paper.

The results of the numerical study show that it can be costly to approximate the inter-arrival times with exponential distributions. Especially in the cases studied when ρ was significantly smaller than 1 and the demand size remained constant and equal to 1, the expected cost increase of assuming exponential inter-arrival times was very high. For several problems this increase was more than 200%. Approximating the Weibull distribution with the gamma distribution worked in general well, generating the same solution in 54 out of 60 problem scenarios.

2.6.5 Paper V – A Note on Solution Procedures for a Class of Two-echelon Inventory Problems

Paper V constitutes a technical note that shows how additional information can be extracted from existing inventory control models. The stochastic multi-echelon literature contains several exact models that analyze the expected costs without first determining the inventory level distribution. These models originate with Axsäter (1990) and are based on similar methodologies. They follow an arbitrary unit as it travels through the system and analyze the expected costs incurring on this unit. The main result of this note is the derivation of the inventory level distribution from the cost function for this group of models. Through the inventory level distribution, several important service measures can be determined, such as the ready rate and fill rate. The note also shows that a well-known relationship from the single-echelon literature between the ready rate and holding and backorder costs also hold for the retailers in the same group of systems.

The presented results are valid for any stock point i where there exists a control variable that fulfills the conditions of an *adjusting control variable*, S_i. These conditions are:

1. There are holding and backorder costs per unit and time unit at stock point i.
2. A shift in S_i by Δ units shifts the long run distribution of the inventory level with Δ units.
3. S_i does not affect any other expected system costs than the holding and backorder costs at stock point i.

The note shows that these conditions are fulfilled by Axsäter (1990), which studies a continuous review base-stock policy system facing Poisson demand. It is also explained how the results can be generalized to other customer demand processes (compound Poisson and compound renewal), other replenishment policies (installation and echelon stock (R,Q) policies) and models using periodic replenishments. Thus, apart from Axsäter (1990) the presented results are valid for the following models, which do not determine the inventory level distributions; Axsäter (1993a,b, 1997, 1998), Forsberg (1995, 1997a,b) and Marklund (2002, 2011), see Section 2.4 for short descriptions of the models. Important prerequisites for the results to hold in the cited models include; central warehouse replenishment policies that work independently of the *adjusting control variable* S_i, FCFS allocations, and partial or complete delivery policies at the central warehouse.

2.7 Contributions

The main contribution of the thesis is that it extends the scope of stochastic multi-echelon inventory systems that can be analyzed exactly. It complements and expands the existing literature primarily by including transportation and delivery decisions and more realistic transportation cost and emission structures to the models, and by allowing more detailed demand information to be considered. Below follow descriptions of the main contributions of Paper I-V separately.

Paper I introduces new aspects when analyzing distribution systems. To the best of our knowledge this paper is the first model studying a stochastic multi-echelon system that includes volume dependent transportation costs. It is also, to our knowledge, the first stochastic multi-echelon model that explicitly considers emissions in the optimization. The main analytical contribution of the paper is the derivation of the distribution of the size of an arbitrary shipment leaving the central warehouse to any retailer group. This distribution enables the analysis of the exact costs for a range of distribution systems with shipment consolidation, where the transportation costs and emissions are dependent on the size of the shipment.

Like Paper I, Paper II generalizes the model of Marklund (2011), but in other directions. Firstly, it extends it to handle compound Poisson demand instead of Poisson demand. It can therefore be applied to systems where the variance to mean ratio is larger than or equal to one. Secondly, it generalizes the analysis to handle fill rate constraints in addition to backorder costs. This extension further improves the practical relevance, as fill rate constraints are commonly used in the industry. Thirdly, it provides a joint optimization procedure for the shipment

intervals, the reorder points and the order-up-to levels, whereas Marklund (2011) determines the shipment interval based on a heuristic.

The main analytical contribution is the cost evaluation, or perhaps, more precisely, the derivation of the distribution of the backorders destined to a specific retailer within this analysis. Also the methodology, with the new concept referred to as *nominal inventory position*, is a contribution, as it has already proven to be useful in other contexts.

Paper III introduces a more realistic cost structure to distribution systems with batch ordering policies. By introducing costs for splitting orders, this paper is able to compare and analyze the choice between dispatching partial or complete orders. Apart from the previously assumed partial and complete ordering policies, we introduce a new Mixed State-Dependent (MSD) policy, which has a performance guarantee over the other two. The main analytical contribution of this paper is the exact derivation of the exact expected costs under this policy. In this analysis, new useful features of the *nominal inventory position* are observed and used.

The numerical study indicates that substantial savings can be made by using the MSD policy compared to the partial or complete delivery policies. It is also worth noting that the performance guarantee, mentioned above, holds for any set of reorder points. In practical applications for large systems, it can therefore be interesting to optimize the reorder points with some faster heuristic and simply use the MSD delivery choice algorithm to decide on how to deliver in each particular instance.

Paper IV presents an exact analysis of a distribution system where the customer demands at the retailers follow compound renewal processes. Thus the customer inter-arrival times may follow any continuous distribution and the amount each customer orders may follow any discrete distribution. To the best of our knowledge, this is the first stochastic multi-echelon inventory distribution system analyzed exactly, where the analysis is not based on exponential customer arrivals. This contribution is strengthened by a numerical study showing that the exponential assumption can generate very poor results.

Paper V presents three results for a group of stock points in multi-echelon systems with *adjusting control variables* defined in Section 2.6.5. The main result is the derivation of the inventory level distribution from the cost function. This result extends the scope of several papers that analyze the costs by following an arbitrary unit as it travels through the system. The second result is a generalization of a relationship between the holding and shortage costs and the ready rate, which is well known for single-echelon systems, to hold for a group of stock points in multi-echelon systems. Finally it proves that the costs are convex in the *adjusting control parameter*.

2.8 Future Research

Based on this thesis, there are several interesting directions for future research. Regarding the shipment consolidation policies studied in Papers I and II, a logical path would be to extend the analysis of the shipment size, transportation costs and emissions in Paper I to the case of compound Poisson demand analyzed in paper II. It would also be interesting to evaluate other transportation cost and emission structures. Furthermore, it would be relevant to compare the results of this time-based shipment consolidation policy with other shipment consolidation policies, for instance quantity-based shipment consolidation. Such a policy consolidates and ships a predetermined quantity (e.g. a truckload) of units for every shipment from the central warehouse. This is interesting, not least from a sustainability perspective, as it assures maximum utilization of transportation capacity.

Based on the research in Paper III, it would be interesting to develop and analyze other delivery policies in batch ordering systems. One development would be to allow the central warehouse to make dynamic decisions between partial and complete deliveries. This means that even if the central warehouse initially expects it to be beneficial to wait for a complete delivery, demands occurring at the retailer can change the circumstances in favor of a partial delivery, triggering a "delayed partial delivery". Another possibility would be to consider a partial delivery policy that cancels the remaining order after the available units have been shipped. Note however, that this policy causes analytical challenges as it alters the inventory position both at the central warehouse and at the retailer. It would also be relevant to extend the analysis of Paper III to handle compound Poisson demand.

Paper IV is, to the best of our knowledge, the first paper analyzing compound renewal demand in a multi-echelon distribution system. It would be interesting to extend the results to other replenishment policies under this demand structure. Possible examples include (R,Q) and (s,S) policies with continuous or periodic replenishments. However, it would also be interesting to see how a delayed ordering policy would perform in a multi-echelon setting. It has been shown in single-echelon settings that delaying orders can create significant cost savings when the customer inter-arrival times are non-exponential, see, for instance Axsäter and Viswanathan (2012) and Syntetos et al. (2016). Generalizing this to multi-echelon distribution systems would be relevant, especially as the times between orders placed to the central warehouse from retailers that batch their customer orders are non-exponential even if their customers arrive with exponential inter-arrival times.

For all the exact procedures presented in Papers I-IV, the computational times become long, for instance, when the amount of retailers or the expected demand per time unit increases. Therefore, finding fast and stable heuristic

solutions to all the presented problems constitutes relevant paths for future research.

References

Andersson, J. 1999. Exact analysis of delivery policies in multi-echelon inventory systems. In *Coordinated multi-stage inventory systems with stochastic demand*. Doctoral Thesis, Lund University.

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

Axsäter, S. 1993a. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* **41** 777-785.

Axsäter, S. 1993b. Optimization of order-up-to-S policies in two-echelon inventory systems with periodic review. *Naval Research Logistics* **40** 245-253.

Axsäter, S. 1997. Simple evaluation of echelon stock (R,Q)-policies for two-level inventory systems. *IIE Trans* **29** 661–669.

Axsäter, S. 1998. Evaluation of installation-stock based (R,Q)-policies for two-level inventory systems with Poisson demand. *Operations Research* **46** S135–S145

Axsäter, S. 2000. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand. *Operations Research* **48** 686-696.

Axsäter, S. 2003. Supply chain operations: Serial and distribution inventory systems. In Graves, S. C., T. de Kok., eds, *Handbooks in operations research and management science, vol 11: Supply chain management: Design, coordination and operation,* Elsevier.

Axsäter, S. 2006. *Inventory control,* second edition. Berlin: Springer.

Axsäter, S., J. Marklund. 2008. Optimal position based warehouse ordering in divergent two-echelon inventory systems. *Operations Research* **56** 976–991.

Axsäter, S., J. Marklund. 2010. *Decision sciences*, Encyclopedia of library and information sciences, third edition **1:1** 1450-1457.

Axsäter, S., S. Viswanathan. 2012. On the value of customer information for an independent supplier in a continuous review inventory system. *European Journal of Operational Research* **221** 340-347.

Chen, F., Y.S. Zheng. 1997. One warehouse multi-retailer systems with centralized stock information. *Operations Research* **45** 275-287.

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

Edgeworth, F. 1888. The mathematical theory of banking. *J. Royal Statist. Soc.* **51** 113–127.

Federgruen, A., P. Zipkin. 1984. Computational issues in an infinite-horizon, multi-echelon inventory model. *Operations research* **32** 818-836.

Forsberg, R. 1995. Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research* **81** 143-153.

Forsberg, R. 1997a. Exact evaluation of (R,Q)-policies for two-level inventory systems with Poisson demand. *European Journal of Operational Research* **96** 130-138.

Forsberg, R. 1997b. Evaluation of (R,Q)-policies for two-level inventory systems with generally distributed customer inter-arrival times. *European Journal of Operational Research* **99** 401-411.

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

Graves, S. C. 1996. A multi-echelon inventory model with fixed replenishment intervals. *Management Science* **42** 1-18.

Gürbüz, M., K. Moinzadeh, Y.P. Zhou. 2007. Coordinated replenishment strategies in inventory/distribution systems. *Management Science* **53** 293-307.

Hillier, F. S., G. J. Lieberman. 2010. *Introduction to operations research*, ninth edition. McGraw-Hill.

Howard, C. 2013. Real-time Allocation Decisions in Multi-echelon Inventory Control. Doctoral Thesis, Lund University.

Howard, C., J. Marklund. 2011. Evaluation of stock allocation policies in a divergent inventory system with shipment consolidation. *European Journal of Operational Research* **211** 298-309.

Iglehart, D. 1963. Optimality of (s,S) policies in the infinite-horizon dynamic inventory problem. *Management Science* **9** 259-267.

Kruse, W. K. 1979. An Exact N Echelon Inventory Model: The Simple Simon Method. Technical Report, March, *U.S. Army Inventory Research Office*, Philadelphia

Marklund, J. 2002. Centralized inventory control in a two-level distribution system with Poisson demand. *Naval Research Logistics* **49** 798-822

Marklund J. 2006. Controlling Inventories in Divergent Supply Chains with Advance-Order Information. *Operations Research* **54** 988-1010.

Marklund, J. 2011. Inventory control in divergent supply chains with time based dispatching and shipment consolidation. *Naval Research Logistics* **58** 59-71.

Marklund, J., K. Rosling. 2012. Lower bounds and heuristics for supply chain stock allocation, *Operations Research* **60** 92-105.

Moinzadeh K. 2002. A Multi-Echelon Inventory System with Information Exchange, *Management Science* **48** 414-426.

O'Connor, P., A. Kleyner. 2012. Practical Reliability Engineering, 5th Edition. New Jersey: John Wiley & Sons.

Paterson, C., G. Kiesmüller, R. Teunter, K. Glazebrook. 2011. Inventory Models with Lateral Transshipments: A Review. *European Journal of Operational Research* **210** 125-136.

Porteus, E. L. 1971. On the optimality of generalized (s,S) policies. *Management Science* **17** 411-427.

Rosling, K. 1989. Optimal inventory policies for assembly systems under random demands. *Operations Research* **19** 565-579.

Shang, K., Z. Tao, S. Zhou. 2015. Optimizing reorder intervals for two-echelon distribution systems with stochastic demand. *Operations Research* **63** 458-475.

Silver, E. A., D. F. Pyke, R. Peterson. 1998. Inventory management and production planning and scheduling, third edition. John Wiley & Sons, New York.

Simon, R.M. 1971. Stationary Properties of a Two Echelon Inventory Model for Low Demand Items. *Operations Research* **19** 761–777.

Stenius, O. 2014. Multi-echelon inventory control with integrated shipment decisions. Licentiate thesis, Lund University.

Syntetos, A.A., R.H. Teunter, M.Z. Babai, S. Tranchel. 2016. On the benefits of delayed ordering. *European Journal of Operational Research*, forthcoming.

Veinott, A. F. 1966. On the optimality of (s,S) inventory policies: New conditions and a new proof. *SIAM Journal of Applied Mathematics* **14** 1067-1083.

Zheng, Y. S. 1991. A simple proof for the optimality of (s,S) policies in infinite-horizon inventory systems. *Journal of Applied Probability* **28** 802-810.

Zipkin, P. H. 2000. *Foundations of inventory management.* McGraw-Hill.

Paper I

Sustainable Multi-echelon Inventory Control with Shipment Consolidation and Volume Dependent Freight Costs

Olof Stenius ● Johan Marklund ● Sven Axsäter

Department of Industrial Management and Logistics, Lund University
christian.howard@iml.lth.se ● olle.stenius@iml.lth.se

Abstract

This paper provides exact analysis of a model for sustainable control of a one-warehouse-N-retailer inventory system with time based shipment consolidation. The model setting is inspired by discussions with industry and involves the possibility to reserve intermodal transportation capacity, in combination with truck transports available on demand. Inventories are reviewed continuously while shipments from the warehouse are consolidated for groups of retailers and dispatched periodically. A key result is the derivation of the probability mass functions for the number of units on each shipment. This allows for realistic volume dependent freight cost structures and emissions to be included in the model. We show how to jointly optimize the reorder levels, shipment intervals and capacity reservation quantities to minimize the total expected costs. Emissions are taken into consideration by use of a side constraint on the total expected emissions or by introducing emissions costs. A numerical example illustrates how the model can be used for evaluating the cost impact of reducing emissions. The analysis is applicable to both single- and multi-item systems.

1. Introduction

Increasing fuel prices and environmental concerns drive a growing interest among companies for more sustainable distribution and freight transportation systems. An important aspect of this challenge is to reduce transportation emissions while minimizing total inventory and transportation costs. In this paper we consider these issues in the context of a distribution system with a central warehouse that replenishes N non-identical retailers (or local warehouses) using a time based shipment consolidation policy with volume dependent freight costs and emissions, and intermodal transport options.

Our research is motivated by discussions with several Swedish companies having one (or a few) central warehouse(s) in central Europe or in Scandinavia, and local warehouses spread across Europe. Spurred by ambitions to reduce total costs and transportation emissions they investigate (or have already embarked on) intermodal transport solutions where goods from the central warehouses are primarily shipped to the local warehouses by train (or in some cases specialized low emission trucks e.g. with extended trailer length and/or alternative fuel engines). This typically means that a transport provider offers a truck-train-truck solution, where (shuttle) trains leave periodically from logistics hubs according to predefined schedules. The periodic shipment schedules facilitate high capacity utilization of the trains, and more reliable transport times. The latter because the transport provider is in a better position to negotiate slot times on the railway systems and avoid transit delays. Procurement of these intermodal transportation services typically requires that the company reserves capacity on the train (or on the specialized trucks) in advance for a given contract period. To assure reliable transport lead-times, which are typically specified in contractual agreements with the shipper, the transport provider complements the intermodal option by regular diesel truck transports. This option is used when the reserved capacity at a given shipment instance is insufficient to transport the entire order volume, or in case of disruptions. Most often the stated goal is that the inventory locations should experience the same transportation lead-time for the entire shipment irrespectively of how individual items are transported.

A fundamental question these companies face is how to leverage this type of intermodal transportation solution to reduce transport emissions while minimizing the total inventory and transportation costs of their distribution systems? The question is challenging because decisions regarding shipment frequencies, consolidation policies and how much train capacity to reserve, are intertwined with inventory decisions at the central warehouse and at the local warehouses.

The model we present addresses these issues and offers means to analyze the tradeoff between transport emissions and total costs, and to optimize system performance. It assumes a centralized

system where the central warehouse has the mandate to control the inventories at the retailers, for example, through a VMI (Vendor Managed Inventory) program. Moreover, the IT systems are integrated to the extent that real time inventory and point-of-sale information for the entire system is available at the warehouse. Motivated by fixed costs for handling and shipping goods from the central warehouse, the retailers are clustered into retailer groups to which consolidated shipments are dispatched periodically from the central warehouse. (The use of periodic shipment schedules in practice, and the advantages it may bring are well documented in the literature, see, for example, Gaur and Fisher (2004).) The time between shipments to a given retailer group is a decision variable referred to as the (constant) shipment interval for the group in question. At each dispatching opportunity the warehouse ships all demanded units that are available using either a primary intermodal transportation option (where by far the longest traveled distance is by train or specialized low emission trucks), or an alternative transportation option (typically regular diesel trucks). Transportation lead-times are assumed to be constant but not identical. For the intermodal option, capacity must be reserved in advance (how much to reserve is a decision variable), and the reserved capacity is used to its fullest extent before the alternative transportation option may be used. The analysis focuses on the single-item case with Poisson demand and First-Come-First-Served (FCFS) allocations, but extensions to multi-item settings are provided.

A key technical contribution of our work is the exact derivation of the probabilities for different shipment quantities to occur under the time based consolidation policy. Knowledge about these probabilities allow for great flexibility in evaluating different types of volume dependent cost structures, for example, with vehicle or load carrier dependent fixed and variable costs. Similarly, it also allows for evaluation of expected emissions associated with different transportation options.

The remainder of this section is devoted to an overview of the related literature. Section 2 provides a detailed model formulation. Section 3 presents the analysis of the expected transportation costs and emissions. Section 4 explains the proposed cost optimization procedures for systems with or without given emissions constraints. An illustrative example is found in Section 5, Section 6 discusses generalizations, and Section 7 concludes.

1.1 Related literature

Our work is closely related to Marklund (2011), which considers the same type of inventory system as we do but for more restrictive transportation cost structure. More precisely, a fixed cost is assumed for every scheduled shipment to each retailer group, regardless of the quantity shipped (even when no units are shipped). The main technical contribution is the derivation of an exact recursive procedure for obtaining the expected inventory holding and backorder costs for all stock points, which

we capitalize on in our present work. In contrast, we contribute with modeling and exact analysis of capacitated volume dependent shipment costs and transportation emissions, allowing for combinations of transportation modes and evaluation of expected emissions. We also provide a procedure for jointly optimizing shipment intervals (to each retailer group), reserved intermodal capacity (to each retailer group), and reorder levels (at all stock points), and show how transportation emissions can be considered in this optimization.

Howard and Marklund (2011) build on Marklund (2011) and investigate by simulation the impact of using state dependent myopic policies instead of FCFS to allocate items to the retailers in a retailer group. For the same system Howard (2013) considers two alternative state dependent allocation policies which are guaranteed to not perform worse than FCFS. The conclusion from these two papers is that overall FCFS performs well, but there may be cost benefits of using state dependent allocation policies, particularly if the allocation decision is postponed to the moment of delivery. Stenius et al. (2015) extend the model in Marklund (2011) to compound Poisson demand. Gürbüz et al. (2007) also study joint inventory and transportation decisions, but in a VMI setting where the central warehouse is a cross-docking facility that is not allowed to hold any stock.

Apart from these articles, our current work is related to the literature on divergent continuous review multi-echelon inventory systems without consolidated shipments. Simon (1971), Graves (1985) and Axsäter (1990) present methods for exact and approximate evaluation of continuous review models with (S-1, S) policies, FCFS allocation and Poisson demand. Generalizations to compound Poisson demand and/or batch ordering are provided in Axsäter (1993a, 1997, 2000), Forsberg (1995, 1997) and Chen & Zheng (1997). The same research stream also encompass papers investigating more general replenishment policies, allocation policies or delivery policies, including Marklund (2002, 2006), Moinzadeh (2002), and Axsäter & Marklund (2008) and Howard & Stenius (2013).

Because of the periodic shipments from the central warehouse, our work is connected to the literature on periodic review multi-echelon inventory control. In contrast to what is assumed in our work, the main body of this literature does not make use of real-time information (see, for instance, Federgruen, 1993, Houtum et al., 1996, Cachon & Fisher, 2000, Axsäter et al., 2002, Özer, 2003, Chu & Shen, 2010, Marklund & Rosling, 2012, and references therein). Inventory levels are observed only when replenishments may be placed. Exceptions include Graves (1996), Axsäter (1993b) and Shang et al. (2014), which assume virtual allocation of orders under Poisson demand arrivals and constant transportation times. This means that demand is monitored continuously and satisfied in a FCFS sequence although replenishment orders (at all sites) can be placed only at preset times according to base-stock policies. Graves (1996) assumes fixed but not necessarily constant replenishment intervals,

and provides exact characterization of the inventory at any time and site in divergent distribution networks with two or more echelons. In these respects our model is more restrictive as it considers periodic shipments in two echelon systems. Axsäter (1993b) focuses on the special case of fixed and nested replenishment intervals and present a fast recursive procedure for evaluating expected holding and backorder costs. In spirit this procedure is similar to the method in Marklund (2011), used in our present work. Shang et al. (2014) consider the replenishment intervals as decision variables and show how to optimize them together with the base stock levels in the system. This is related to the optimization of the shipment intervals in our study. Volume dependent freight costs and transport emissions are not considered in any of the papers mentioned above, which sets our work apart. Another difference is that in our model the central warehouse uses a continuous review (R,Q) policy.

Our work is also associated with the shipment consolidation literature focusing on single-echelon settings (e.g. Çetinkaya & Lee, 2000, Axsäter, 2001, Çetinkaya & Bookbinder, 2003, Chen et al., 2005, Çetinkaya et al., 2008, Mutlu et al., 2010, and Kaya et al., 2012). These papers typically study a vendor, which through VMI contracts decides replenishments (often with negligible replenishment lead-times) and dispatches consolidated shipments to a number of retailers. Similar to our present work Çetinkaya & Bookbinder (2003), Mutlu et al. (2010) and Kaya et al. (2012) allow for volume dependent dispatching costs. Key differences between this stream of literature and our work are that we study a multi-echelon system and explicitly consider emissions in the optimization.

Disregarding the shipment consolidation aspect, Tempelmeier & Bantel (2015) approximate the probability distribution of the daily transportation volume from an inventory location using a periodic review (R,Q) policy. The daily volume is dispatched using limited in-house transportation capacity extended by a more costly external option. The authors advocate the importance of jointly optimizing inventory and transportation decisions and show that reducing the safety stock increases the variability of the daily transportation volume and the associated costs. Our approach is different, providing exact analysis and optimization of inventory and transportation decisions with respect to both costs and emissions in multi-echelon settings with shipment consolidation.

The research on sustainable supply chain management has increased significantly during the last decade. Until recently, most of the quantitative models have been focusing on closed-loop reverse logistics systems, or waste management (see, for example, Kleindorfer et al. 2005, Corbett and Klassen, 2006, Srivastava, 2007, and Dekker et al., 2012, for overviews). Lately, the interest in green inventory management models where emissions are considered has increased. For an overview of this literature, which so far is dominated by deterministic lot sizing models and newsvendor type models,

we refer to Marklund & Berling (2015). As far as we know, our work is the first to consider volume dependent shipment costs and transport emissions in a stochastic multi-echelon inventory setting.

2. Model formulation

Our model includes a central warehouse that replenishes N non-identical retailers facing independent Poisson demand of a single item (extension to multi-item systems is available in Appendix C). The central warehouse has access to real-time inventory information about the entire system, and replenishes its stock from an outside supplier/manufacturer with constant lead-time, L_0. Moreover, as indicated above, complete backordering and FCFS allocation is assumed at all stock points.

The retailers are divided into K retailer groups (K ≤ N) and there are N_k retailers belonging to retailer group k, $\sum_{k=1}^{K} N_k = N$. The set of retailers belonging to retailer group k is denoted, Ω_k. The shipments from the central warehouse to each retailer group are consolidated, offering a potential for reducing transportation costs and emissions. More precisely, the central warehouse dispatches a shipment to all retailers within retailer group k (1 ≤ k ≤ K) every T_k time units. The shipment interval, T_k, is assumed to be a positive multiple of some smallest shipment interval T_{min}. The retailer groups are taken as given inputs to the model. They may, for example, be determined by geographical proximity or by use of a vehicle routing method (see, for example, Toth and Vigos, 2001 for an overview). Clearly, the configuration of the retailer groups can affect the performance of the system, and the presented model can be used for evaluating different alternatives with regards to expected costs and emissions. However, optimizing the configuration of retailer groups involve many issues beyond the scope of our present model and is left for future research.

To each of the retailer groups, shipments are made either with a primary intermodal transportation option, where capacity is reserved in advance, or with an alternative transportation option where capacity is unlimited. The latter is used only if a shipment quantity exceeds the reserved intermodal capacity. The set of possible capacity reservations for the primary option (expressed in number of units) to retailer group k is denoted W_k. For each retailer group there is a possibility not to reserve any capacity, $w_k=0$ (i.e., $0 \in W_k$ ∀k=1,2,…K), and to use only the alternative option. In practice, the capacity reservation opportunities to a specific retailer group may include the possibilities to reserve nothing or an integer multiple of: whole trains, freight cars, containers or load carriers. Based on our discussion with industry, where reliability in delivery times is emphasized, the main part of the analysis assumes that the transportation lead-times from the central warehouse to retailer i (including picking at the central warehouse, loading, transporting, unloading at retailer i etc.), L_i, is constant and independent of the transportation option used. An exact analysis of the more general case

of different transportation lead-times for the primary and alternative options is available in Appendix D. In either case, the replenishment lead-time for retailer i (i.e., the time from order placement at the central warehouse until the unit is available at retailer i) is stochastic.

The central warehouse uses an (R_0, Q_0) replenishment policy, meaning that an order of Q_0 units is placed every time the inventory position (defined below) reaches R_0. R_0 is a decision variable while the order quantity Q_0 is presumed to be given by the outside supplier/manufacturer, taking its production set up costs etc. into consideration. Q_0 may also be determined by a deterministic EOQ method, as suggested in Zheng 1992 and Axsäter 1996. Even though Q_0 is not a decision variable in our model, our method can of course be used repeatedly to evaluate different Q_0 options.

As indicated above, the warehouse has access to real time inventory and point-of-sale information from all stock points. This means that as soon as a demand occurs at a retailer, the information is transferred to the central warehouse. In effect this means that all retailers apply (S_i-1, S_i) ordering policies. However, it is important to note that replenishments are typically not delivered one unit at a time. Because of the periodic shipment policy used at the warehouse replenishments are consolidated by optimizing the shipment intervals. One can note that as a consequence of the system set up, the aggregated demand at the warehouse is also Poisson.

Focusing on the inventory process at the central warehouse, the FCFS allocation implies that a unit will be reserved for retailer i at the moment this retailer experiences a customer demand. This reserved unit may either be available at the central warehouse or not. In case the unit is available, it becomes *qualified* for shipment and will be part of the next shipment to retailer i. In case the reserved unit is not yet available, it will be backordered at the central warehouse. The unit will then become *qualified* for shipment (and the backorder cleared) at the moment it is delivered to the central warehouse from the outside supplier. It will then be sent on the next scheduled shipment to retailer i. Emergency shipments of units outside the periodic schedules are not allowed. The motivation stems from our industry discussions, where articulated ambitions were to reduce transport emissions by shipment consolidation and use of intermodal shuttle train solutions. Introducing the possibility of emergency shipments, for example, shipping backordered units that have missed a scheduled shipment immediately upon their arrival to the central warehouse is an interesting direction for future research.

The inventory level at the central warehouse at time t_0, $IL_0(t_0)$, is defined as the number of available units minus the amount of backordered units (the qualified units awaiting shipment are not included as they are already reserved for delivery to specific retailers). Equivalently, the inventory position at time t_0, $IP_0(t_0)$, is defined as the inventory level $IL_0(t_0)$, plus all outstanding orders.

There are holding costs at all stock points, backorder costs at all retailers, and fixed and variable shipment costs for replenishments sent from the central warehouse.

$\mathbf{T}$ = vector of shipment intervals to all retailer groups = $\{T_1,\ldots,T_K\}$, $T_k = n_k \cdot T_{min}$ $\forall$ k=1,2,...K, where n_k is an integer greater than zero.

$\mathbf{w}$ = vector of capacity reservation quantities to all retailer groups = $\{w_1,\ldots,w_K\}$, $w_k \in W_k$ $\forall$ k.

$\mathbf{S}$ = vector of the order-up-to levels at all retailers = $\{S_1, \ldots, S_N\}$.

λ_i = expected demand per time unit at retailer i.

λ_0 = expected total demand per time unit at the central warehouse = $\sum_{i=1}^{N} \lambda_i$.

$\lambda_{(k)}$ = expected demand per time unit at the central warehouse from retailer group k, = $\sum_{i \in \Omega_k} \lambda_i$.

h_i = holding cost per unit and time unit at stock point i, i = 0,1,...,N

b_i = backorder cost per unit and time unit at retailer i, i =1,2,...,N

x^+ = max(0,x) and analogously, $x^- = $ max(0,-x)

$TC(R_0,\mathbf{S},\mathbf{T},\mathbf{w})$ = expected total cost per time unit

$TIC(R_0,\mathbf{S},\mathbf{T})$ = expected holding and backorder cost per time unit

$TSC(R_0,\mathbf{T},\mathbf{w})$ = expected shipment cost per time unit

Note that when the primary and alternative options offer the same transportation lead-time, the expected inventory holding and backorder cost per time unit, TIC, is unaffected by the reserved capacity $\mathbf{w}$. Also note that the expected shipment cost per time unit, TSC, is unaffected by the base-stock levels, $\mathbf{S}$, (see Section 3 for explanation), and recall that Q_0 is given. Hence,

$$TC(R_0,\mathbf{S},\mathbf{T},\mathbf{w}) = TIC(R_0,\mathbf{S},\mathbf{T}) + TSC(R_0,\mathbf{T},\mathbf{w}). \tag{1}$$

Our focus is to evaluate the shipment costs, $TSC(R_0,\mathbf{T},\mathbf{w})$, and the emissions in the system. The inventory costs, $TIC(R_0,\mathbf{S},\mathbf{T})$, can be obtained by the recursive method presented in Marklund (2011).

2.1 Shipment costs and emissions

The considered shipment costs consist of three parts; (i) fixed costs for the reserved transportation capacity on the primary option for each scheduled shipment (may include fixed costs for picking, receiving, administration, etc.), (ii) fixed costs for each load carrier used on the alternative option, and (iii) linear costs per unit shipped on each of the two transport options. We define:

$\alpha_k'(w_k)$ = fixed cost for each scheduled shipment to retailer group k when a capacity of $w_k \in W_k$ is reserved on the primary transportation option

A_k = number of units on a single load carrier for the alternative option to retailer group k.

α_k'' = fixed cost per load carrier on the alternative option for retailer group k

c_i' = variable cost per unit shipped by the primary option to retailer i

$c_i'' =$ variable cost per unit shipped by the alternative option to retailer i

$\Delta c_i =$ variable cost increase per unit for shipping with the alternative option to retailer i $= c_i'' - c_i'$

Because all units are shipped on one of the two transport options, only the cost increase per unit for shipping with the alternative option, Δc_i, matters in the analysis. For exposition reasons we assume that this cost increase is equal for all retailers within each retailer group, and define $\Delta c_{(k)} = \Delta c_i$ $\forall i \in \Omega_k$. Relaxation of this assumption is discussed in Section 6.

The fixed cost $\alpha_k'(w_k)$ (i.e. capacity reservation, picking, receiving, administrative costs etc.) is incurred regardless of the number of units actually shipped. However, as we derive the probabilities of different shipment quantities, it is easy to modify the analysis so that only part (or none) of these costs are incurred in situations where no units are shipped to retailer group k.

The freight transportation emissions of greenhouse gases are included in the model by a combination of fixed and variable emissions parameters.

$\beta_k'(w_k) =$ fixed emissions for each scheduled shipment to retailer group k when a transportation capacity of $w_k \in W_k$ is reserved on the primary option

$\beta_k'' =$ fixed emissions per load carrier or vehicle used on the alternative option to retailer group k

$e_i' =$ emissions per unit shipped to retailer i by the primary option

$e_i'' =$ emissions per unit shipped to retailer i by the alternative option

$\Delta e_i =$ emissions increase per unit shipped to retailer i by the alternative option $= e_i'' - e_i'$

$TE(R_0, \mathbf{T}, \mathbf{w}) =$ Expected total emissions per time unit for the system

The fixed emissions $\beta_k'(w_k)$ for the primary option, are incurred regardless of how much of the reserved capacity that is actually used by the company. The motivation for this is that the shuttle trains (or specialized low emissions truck) will cause emissions when they run whether the reserved capacity is utilized or not. If instead, we assume that the unused capacity is sold to somebody else, this can be modeled by setting $\beta_k'(w_k) = 0$ $\forall w_k \in W_k$. Analogously to the cost analysis, the emissions increase per unit for shipping with the alternative option, Δe_i, is used in the analysis instead of e_i' and e_i''. We also assume that $\Delta e_{(k)} = \Delta e_i$ $\forall i \in \Omega_k$ and discuss the relaxation of this assumption in Section 6.

3. Analysis

In this section we derive the expected shipment cost per time unit and the expected emissions per time unit for given values of R_0, $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{w}$. The analysis is based on deriving the probability mass function (pmf) of the shipment quantities to each retailer group k (k=1,…,K). Letting t_0 be the time at which a shipment leaves for retailer group k, we define:

$M_{(k)}(t_0)$ = Shipment quantity to retailer group k (i.e. number of units shipped to all retailers in retailer group k at t_0), $M_{(k)} \equiv M_{(k)}(t_0)$

Section 3.1 derives the pmf, $P(M_{(k)} = m_{(k)})$. Based on this analysis we obtain the expected shipment costs and emissions in Section 3.2. All proofs are deferred to Appendix A.

3.1 The shipment quantity

A key step in the analysis is to determine the probability mass function (pmf) of M, the total amount of units that has been qualified for shipment (to all retailers in all retailer groups) since the last shipment departed. The pmf of the shipment quantity to each retailer group is then obtained by binomial disaggregation. This disaggregation technique is similar to the one used in Simon (1971) and Graves (1985) for determining the number of backordered units destined to each retailer in a divergent inventory system with (S-1,S) policies and Poisson demand. Let,

$$Bin(a,b,p) = \binom{b}{a} p^a (1-p)^{b-a}.$$

Proposition 1: The probability of shipment quantity $m_{(k)}$ to retailer group k at time t_0 can be determined through binomial disaggregation of M, the total number of units qualified for shipment to all retailers in all retailer groups during time period $[t_0 - T_k, t_0)$,

$$P\{M_{(k)} = m_{(k)}\} = \sum_{m=m_{(k)}}^{\infty} P\{M = m\} Bin\left(m_{(k)}, m, \frac{\lambda_{(k)}}{\lambda_0}\right). \tag{2}$$

To obtain the probability mass function of M, i.e. the number of units becoming qualified for shipment in time interval $[t_0 - T_k, t_0)$, we define:

$D_0(t_1,t_0)$ = demand at the central warehouse in time interval $[t_1,t_0)$, $t_0 \geq t_1$, Poisson distributed with mean $\lambda_0(t_0 - t_1)$.

$\underset{(R_0,Q_0)}{mod}(x)$ = $x + nQ_0$, where n is an integer such that $R_0 < x + nQ_0 \leq R_0 + Q_0$

$IL_0^-(t)$ = number of backordered units at the central warehouse at time t (i.e., reserved units that have been demanded but are not yet qualified for shipment at time t).

The analysis is divided in two cases; (1) $L_0 \leq T_k$ in Section 3.1.1, and (2) $L_0 > T_k$ in Section 3.1.2. However, we first establish Lemma 1 and Lemma 2.

Lemma 1: The number of units qualified for shipment in time interval $[t_1, t_0)$, M, is

$$M = D_0(t_1, t_0) + IL_0^-(t_1) - IL_0^-(t_0). \tag{3}$$

Lemma 2. The inventory position at time $t \geq t_1$ can be obtained as

$$IP_0(t) = \underset{(R_0,Q_0)}{mod}\left(IP_0(t_1) - D_0(t_1,t)\right) \tag{4}$$

9

3.1.1 Analysis of case 1: $L_0 \leq T_k$

When studying the amount of units becoming qualified for shipment between $t_1 = t_0 - T_k$ and t_0 there are two other important points in time (see Figure 1):

$t_0 - L_0 =$ last point in time when orders placed by the warehouse will arrive on time to be shipped at t_0.

$t_1 - L_0 =$ last point in time when orders placed by the warehouse will arrive on time to be shipped at t_1.

These four time instances define three time intervals (A, B and C). The demand at the warehouse in these time intervals are $D_A = D_0(t_1 - L_0, t_1)$, $D_B = D_0(t_1, t_0 - L_0)$ and $D_C = D_0(t_0 - L_0, t_0)$, see Figure 1.

We start the analysis from $t_1 - L_0$, with an inventory position of $IP_0(t_1 - L_0) = x$. This inventory position is uniformly distributed on $[R_0+1, R_0+Q_0]$ (see, Axsäter, 1998 or Marklund, 2002). In order to determine the number of units qualified for shipment in time interval $[t_1, t_0)$ the inventory levels at times t_1 and t_0 need to be obtained. The inventory level at time t_1 can be expressed as

$$IL_0(t_1) = x - D_A . \tag{5}$$

Similarly, the inventory level at time t_0 is

$$IL_0(t_0) = IP_0(t_0 - L_0) - D_C \tag{6}$$

Using (4), the inventory position at time $t_0 - L_0$ follows from the inventory position x at time $t_1 - L_0$.

$$IP_0(t_0 - L_0) = \underset{(R_0, Q_0)}{mod}(x - D_A - D_B) \tag{7}$$
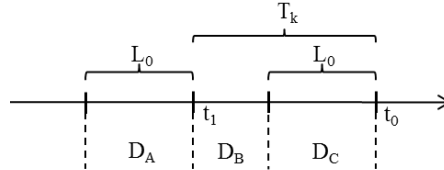
Figure 1. Time intervals when $L_0 \leq T_k$.

The number of qualified units in time interval $[t_1, t_0)$, M, now follows from Lemma 1

$$M = D_B + D_C + (x - D_A)^- - \left(\underset{(R_0, Q_0)}{mod}(x - D_A - D_B) - D_C\right)^- . \tag{8}$$

As x, D_A, D_B and D_C are independent stochastic variables, the pmf of M can in principle be obtained from (8) by convolutions. It is efficient to use the following procedure. Initially let the probabilities for all possible outcomes of M be zero. For each value of x, D_A, D_B and D_C, the resulting value of M can be uniquely computed from (8). Moreover, the probability for this combination of independent outcomes follows directly as x is uniform on $[R_0 + 1, R_0 + Q_0]$ while D_A, D_B and D_C are Poisson distributed. By successively considering new combinations of x, D_A, D_B and D_C we can augment the probability mass for the corresponding M value and eventually obtain the correct pmf. New combinations are obtained by considering larger values of D_A, D_B and D_C (i.e., larger demands).

10

We can disregard higher demands and stop searching new combinations of x, D_A, D_B and D_C when the total probability of M is sufficiently close to 1.

3.1.2 Analysis of case 2: $L_0 > T_k$

When $L_0 > T_k$, there are again four critical points in time to consider in order to determine the number of units qualified for shipment in time interval $[t_1,t_0)$. As before these times are t_0, $t_1 = t_0 - T_k$, $t_0 - L_0$ and $t_1 - L_0$ (see Figure 2). The difference from case 1 is that t_1 now occurs after $t_0 - L_0$. Once again these four time instances define three time periods (D, E, and F), and the demand at the warehouse in these intervals are $D_D = D_0(t_1 - L_0, t_0 - L_0)$, $D_E = D_0(t_0 - L_0, t_1)$ and $D_F = D_0(t_1, t_0)$.

With an initial inventory position x at time $t_1 - L_0$, the inventory level at time t_1 is

$$IL_0(t_1) = x - D_D - D_E. \tag{9}$$

From Lemma 2, $IP_0(t_0 - L_0) = \underset{(R_0, Q_0)}{\text{mod}}(x - D_D)$, and the inventory level at time t_0 is

$$IL_0(t_0) = \underset{(R_0, Q_0)}{\text{mod}}(x - D_D) - D_E - D_F. \tag{10}$$

M can, from Lemma 1, be obtained as

$$M = D_F + IL_0^-(t_1) - IL_0^-(t_0). \tag{11}$$
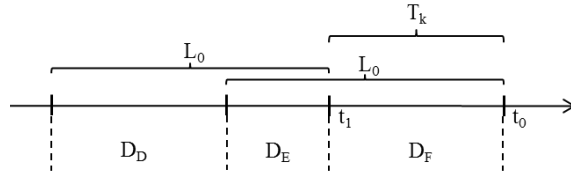
Figure 2. Time intervals when $L_0 > T_k$.

Note again that x, D_D, D_E and D_F are independent random variables. The pmf of M can therefore be determined in the same way as in Section 3.1.1, but with (11) defining the value of M instead of (8).

3.2 Analysis of shipment costs and emissions

We now turn our attention to the shipment costs and the emissions of the system. Let:

$SC_k(w_k, m_{(k)}) =$ shipment costs for a shipment leaving to retailer group k, given the reserved primary capacity, w_k, and the shipment quantity $m_{(k)}$

Based on the analysis of the shipment quantities in §3.1, the expected shipment cost per time unit to retailer group k, TSC_k, is obtained from (12). Note, that this expression focuses on the scheduled shipments, but fully accounts for the possibilities that there may be no units to ship ($m_{(k)}=0$). Also, recall that the cost per unit shipped by the primary transportation option to retailer i, c_i', is excluded from the analysis. To include these costs, the term $\sum_{i \in \Omega_k} \lambda_i c_i'$ should be added to TSC_k in (12)

$$TSC_k = \frac{1}{T_k} \sum_{m_{(k)}=0}^{\infty} P\big(M_{(k)} = m_{(k)}\big) SC_k\big(w_k, m_{(k)}\big). \tag{12}$$

The total expected shipment cost per time unit follows directly

$$TSC\big(R_0, \mathbf{T}, \mathbf{w}\big) = \sum_{k=1}^{K} TSC_k. \tag{13}$$

Given w_k and $m_{(k)}$, the number of units shipped with the alternative transportation mode is $\big(m_{(k)} - w_k\big)^+$. The shipment cost is thus

$$SC_k\big(w_k, m_{(k)}\big) = \alpha_k{}'\big(w_k\big) + n_k \alpha_k{}'' + \big(m_{(k)} - w_k\big)^+ \Delta c_{(k)}, \tag{14}$$

where n_k is the integer that satisfies $\big(n_k - 1\big)A_k < \big(m_{(k)} - w_k\big)^+ \leq n_k A_k$.

The expected emissions are obtained analogously. If $E_k(w_k, m_{(k)})$ denotes the emissions for a shipment with reserved capacity w_k, and a given shipment quantity to retailer group k, $m_{(k)}$, we have

$$E_k\big(w_k, m_{(k)}\big) = \beta_k{}'\big(w_k\big) + n_k \beta_k{}'' + \big(m_{(k)} - w_k\big)^+ \Delta e_{(k)}, \tag{15}$$

where n_k is defined as in (14). The expected emissions per time unit for shipments to retailer group k, TE_k, are now obtained from (16) and the total expected emissions per time unit from (17). Recall that the emissions per unit shipped by the primary transportation option to retailer i, $e_i{}'$, are excluded from the analysis. To include them simply add the term $\sum_{i \in \Omega_k} \lambda_i e_i'$ to the expression for TE_k in (16).

$$TE_k = \frac{1}{T_k} \sum_{m_{(k)}=0}^{\infty} P\big(M_{(k)} = m_{(k)}\big) E_k\big(w_k, m_{(k)}\big). \tag{16}$$

$$TE\big(R_0, \mathbf{T}, \mathbf{w}\big) = \sum_{k=1}^{K} TE_k. \tag{17}$$

Note, as the shipment quantities (in steady state) are independent of the retailer order-up-to levels, so are the transportation costs and the emissions of the system.

4. Optimization

This section presents an optimization method for finding cost optimal values for R_0, $\mathbf{S}$, $\mathbf{T}$, $\mathbf{w}$ for three different scenarios; (i) emissions are not considered, (ii) there is a known per unit cost for the emissions, ρ, and (iii) there is a restriction on the expected emissions per time unit allowed in the system, θ. The optimization method focuses on scenario (iii), as (i) and (ii) are special cases where θ is infinite for (i), and for (ii) the cost parameters include the emissions related costs.

By example it can be shown that the total cost function is not jointly convex in R_0, $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{w}$. However, for given values of R_0, $\mathbf{T}$ and $\mathbf{w}$ the total cost is separable and convex in the retailer order up

to levels $\{S_1,S_2,\ldots,S_N\}$. This follows from Marklund (2011) because the emissions and transportation costs associated with **w** are unaffected by **S** when R_0 and **T** are fixed.

The optimization procedure can be described as a bounded search over all relevant combinations of **T**, R_0 and **w** ($w_k \in W_k \; \forall\, k=1,\ldots,K$), using the convexity for optimizing **S** (given **T** and R_0). More precisely, for each relevant combination of **T**, **w** and R_0, investigate whether the solution is environmentally feasible (if $TE(R_0,\mathbf{T},\mathbf{w}) < \theta$). If so, find the optimal order-up-to level for each retailer i, S_i, using that the total cost is separable and convex in the retailer order up to levels **S**. If the total expected cost is lower than $\overline{TC}$, the lowest cost so far, $\overline{TC}$ is updated and the associated solution is saved. When the search is finished, the optimal solution for given the emissions constraint, $TC^*(\theta)$, has been found and the minimum total expected cost is $\overline{TC}$. The relevant search space is determined by upper and lower bounds on the optimal shipment intervals (i.e., $T_k^l \leq T_k \leq T_k^u$ for all k=1,2,…K), by upper and lower bounds on the optimal warehouse reorder point, R_0 (i.e., $R_0^l \leq R_0 \leq R_0^u$), and the finite set of possible capacity reservations, $w_k \in W_k \; \forall\, k=1,\ldots,K$.

A critical step in the optimization procedure is to determine bounds on the shipment intervals and warehouse reorder point. Starting with the latter, we conclude that the results in Marklund (2011), can be transferred to our system. The starting point for the recursive cost evaluation procedure in this paper (and an upper bound on R_0) is $R_0^u = \min\{R_0: P(D(0,L_0) > R_0) < \varepsilon\}$, where ε is a small positive value close to zero. This ensures that an increase in R_0 above R_0^u will not affect the delivery delays caused by backorders at the central warehouse, as they are already (close to) zero. Therefore the inventory levels at the retailers and the shipment quantities will not be affected by further increasing R_0. With analogous reasoning the lower bound for R_0 is set to $R_0^l = -Q_0$ (see Marklund 2011 and Axsäter 1998). It is not difficult to find tighter upper bounds on R_0 (see for example Stenius et al. 2015). However, since the fast recursive procedure in Marklund (2011) is initiated at R_0^u, and in order to determine the costs for R_0^l, the cost optimal solution for all intermediate R_0 values are obtained for free, a tighter upper bound does not improve the computational performance.

For the shipment intervals, an obvious lower bound is $T_k^l = T_{min}$, $\forall\, k=1,\ldots,K$, and an upper bound is provided in Proposition 2.

Proposition 2. All systems with a shipment interval for retailer group k, T_k, larger than

$$T_k^u = \frac{2\overline{TC}}{\lambda_{(k)}h_0} \tag{18}$$

will have an expected total cost that is larger than $\overline{TC}$.

Proposition 2 provides upper bounds on T_k for all k, provided that a feasible solution has been found, i.,e., $\overline{TC} < \infty$. (in scenario (i) and (ii) all solutions are feasible). If θ is set too low, solutions fulfilling the emissions constraint may not exist. To determine if feasible solutions exist one may compute the lowest possible expected emissions per time unit, TE^e. If $TE^e > \theta$, no feasible solution exists. If $TE^e \leq \theta$, a feasible solution can be found by optimizing S for R_0, T and w corresponding to TE^e. For reasons of exposition further analysis of this special case is deferred to Appendix C.

5. Numerical example

This section presents a small example of an inventory system distributing a bulky and expensive item. The objectives are to demonstrate how the model can be used for optimizing the system performance (minimizing expected costs subject to an emissions constraint), and how it can be used to understand the cost impact of enforcing emissions targets. Cost parameters are motivated by discussions with industry and emissions calculations are based on information from the NTM database (NTM, 2011).

The inventory system consists of a central warehouse supplying three retailers in two retailer groups (N=3, K=2). One retailer group consists of two retailers ($\Omega_1 = [1,2]$), and a single retailer constitutes the second retailer group ($\Omega_2 = [3]$). To both groups there exist opportunities to reserve capacity on an intermodal shuttle train solution (primary option), $W_1 = W_2 = [0,5,10,15,20]$ units. The alternative option for both retailer groups is to use direct truck transports, $A_1 = A_2 = 5$ units (i.e., for both options a filled load carrier contains 5 units). The fixed cost per load carrier is 1500 € for the primary option and 2000 € for alternative option. There are also variable transportation costs per unit shipped that are 20 € higher for the alternative option. The emissions are also divided into fixed and variable components. Transporting an empty load carrier by train causes emissions of 200 kg CO_2 equivalents, and transporting it by truck causes emissions of 850 kg CO_2 equivalents. The variable emissions per unit shipped are 10 kg CO_2 equivalents higher for the alternative option. The remaining problem parameters are specified in Table 1.

Table 2 illustrates the results for three problem instances; (I) no emissions constraint exists, (II) The emissions target $\theta = 100$ kg CO_2 equivalents, and (III) $\theta = TE^e$. The cost optimal solution without the emissions constraint provides an expected total cost of 881.89 € per day (obtained from expression (1)), where TSC=629.38 and TIC=252.51). The corresponding expected emissions are 131.67 kg CO_2 equivalents per day (from (17)). Enforcing the emission constraint of $\theta = 100$ kg CO_2 equivalents per day (24.1% lower than the emissions in the cost optimal system without any emissions constraints), leads to a decrease of the transportation costs to 593.44 but an increase of the inventory costs to 319.73, the resulting expected total costs increase with 3.5% to 913.17 € per day. Table 2 shows that more capacity is now reserved on the train (**w** increases from [10,5] to [15,10]), the shipment intervals

(**T**) are longer (to ensure that the reserved capacity is better utilized), and the order-up-to levels at the retailers (**S**) are increased (to balance holding and backorder costs for longer replenishment times). The system with the lowest possible expected emissions ($\theta = TE^e$), reserves maximum capacity on the train, and has even larger **T** and **S** values. For this system, the emissions are 34.5% lower than for the unrestricted cost optimal solution but the expected total costs are 17.5% higher.

Table 1. Problem parameters

Cost Parameters (€)	Emission parameters (kg CO₂ equivalents)
$\alpha_1'(W_1) = \alpha_2'(W_2) = [0,1500,3000,4500,6000]$	$\beta_1'(W_1) = \beta_2'(W_2) = [0,200,400,600,800]$
$\alpha_1'' = \alpha_2'' = 2000$	$\beta_1'' = \beta_2'' = 850,$
$\Delta c_{(1)} = \Delta c_{(2)} = 20$	$\Delta e_{(1)} = \Delta e_{(2)} = 10$
$h_0 = h_i = 10$ €/day, for all i	*Other parameters*
$b_i = 100$ €/day, for all i	$L_0=10$ days, $L_1= L_2 =L_3=2$ days, $T_{min} = 1$ day
	$Q_0=10$ units, $\lambda_1=\lambda_2=\lambda_3=0.5$ units/day

Table 2. Optimal solutions for different θ values.

θ	R_0	**T**	**w**	**S**	TIC^*	TSC^*	$TC^*(\theta)$	TE
∞	10	[10,9]	[10,5]	[8,8,7]	252.51	629.38	881.89	131.67
100	9	[13,17]	[15,10]	[9,9,11]	319.73	593.44	913.17	99.91
TE^e	14	[16,32]	[20,20]	[10,10,17]	447.09	588.95	1036.04	86.29

Table 3 provides further analysis of shipment solutions for the three problem instances. Without an emissions constraint, the train option is used for shipping 87% (86%) of the total number of units delivered to retailer group 1 (retailer group 2), and truck is used for the remaining 13% (14%). When the emissions constraint is tightened to $\theta= TE^e$, 98% of the units are shipped by train, leaving only 2% to be transported by truck. A closer look at the truck option, reveals that the probability of not needing any trucks when a shipment leaves for retailer group 1 and 2, increases from 0.593 and 0.706, respectively, when there is no emissions constraint, to about 87% when $\theta=TE^e$. (These probabilities are obtained directly from the pmf of the shipment quantity $M_{(k)}$ in (2)). We can also see that the utilization of the reserved train capacity is affected. For retailer group 1, it decreases from 87% first to 82% ($\theta=100$), and then to 78% ($\theta=TE^e$) as the emissions constraint is tightened. Intuitively, this is a result of hedging against having to use the truck option. However, for retailer group 2, the capacity utilization first increases from 77% to 79% ($\theta=100$) and then decreases to 78% ($\theta=TE^e$). Thus, the behavior is difficult to predict in general due to the integrality of the decision variables.

Figure 3 provides the lowest expected costs, $TC^*(\theta)$, for all θ values between TE^e and the emissions in the cost optimal solution without the emissions constraint (higher θ values will generate the same solution). It is noteworthy that moderate reductions of emissions can be achieved with only small increases in expected costs. For instance, a system with emissions 16.5% lower than in the cost optimal solution only has 1.5% higher costs. However, for large emissions cuts, the cost increase escalates. Hence, the model can be used for evaluating the cost impact of managerial decisions to

reduce emissions, and for analyzing tradeoffs between emissions and costs. It also prescribes how to achieve the emission targets at minimum cost by choosing optimal values of the decision variables.

Table 3. Analysis of the resulting shipment solutions for different θ values.

θ	Ret grp	Shipment volume		Utlization	Probability of using x trucks				
		train	truck	of w_1, w_2	x=0	x=1	x=2	x=3	x=4
∞	1	87%	13%	87%	0.593	0.360	0.043	0.004	0.000
	2	86%	14%	77%	0.706	0.286	0.008	0.000	0.000
100	1	95%	5%	82%	0.784	0.178	0.036	0.002	0.000
	2	93%	7%	79%	0.763	0.222	0.015	0.000	0.000
TE^e	1	98%	2%	78%	0.876	0.113	0.010	0.000	0.000
	2	98%	2%	78%	0.870	0.118	0.012	0.001	0.000
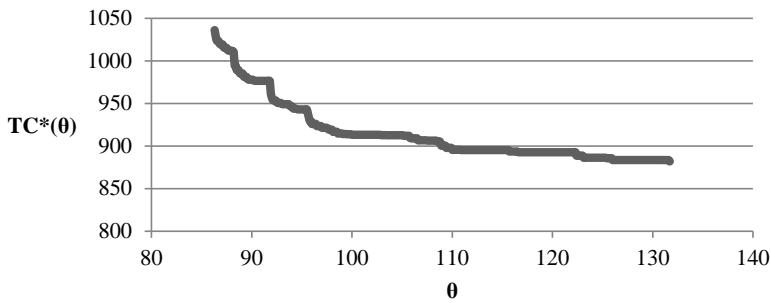
Figure 3. Expected Total Cost for varying θ values.

6. Extensions and generalizations

In practice, shipment consolidation usually involves many different items that are shipped together, emphasizing the relevance of a multi-item perspective. Extending the single-item model to multi-item systems is relatively straightforward. The key is to require that all items with consolidated shipments to retailer group k use the same shipment interval, T_k, and to consider the total costs and emissions across all items and retailers in the system. For a given set of shipment intervals, the inventory costs can then be obtained individually for each item. An added complication is that transportation capacity becomes ambiguous to measure in number of units. Instead a common measure like weight or volume needs to be used. The shipment size distribution (in weight or volume), and thereby the transportation costs and emissions, can be achieved by convolution of the shipment size distributions of the individual items. For the analytical details we refer to Appendix C.

The assumption of equal transportation lead-times for the primary and alternative transportation options may sometimes be questionable. For example, direct truck transports from a point of departure to a final destination is often faster than intermodal train transports. However, the opposite is also possible; reservation of transport capacity in advance can enable efficient flows of goods not achievable with a flexible transportation mode. Appendix D extends the model with an exact analysis

16

of situations where the transportation lead-times times for the primary and alternative transportation options differ. Another assumption used so far is that the cost and the emissions increase per unit for the alternative option over the primary option are equal for all retailers in a retailer group, $\Delta c_{(k)} = \Delta c_i$ and $\Delta e_{(k)} = \Delta e_i \ \forall i \in \Omega_k$. This assumption can be relaxed while maintaining the FCFS principle, letting units demanded first be transported with the primary option. Without loss of generality, we renumber the retailers in retailer group k $\{1,2,\ldots,N_k\}$, and the rest of the retailers from $N_k + 1$ to N. The probability that an arbitrary unit, shipped with the alternative option to retailer group k, is going to retailer i ($i \in \{1,\ldots,N_k\}$) is $\lambda_i/\lambda_{(k)}$. This follows as the order process to the warehouse is Poisson, and the replenishment and consolidation processes at the warehouse (for given system parameters) are unaffected of which retailer within a group the order originates with. The expected cost increase for shipping an arbitrary unit with the alternative option can thus be obtained as

$$\Delta c_{(k)} = \sum_{i=1}^{N_k} \frac{\lambda_i}{\lambda_{(k)}} \Delta c_i. \tag{19}$$

Analogously, the expected emissions increase per unit with the alternative option is

$$\Delta e_{(k)} = \sum_{i=1}^{N_k} \frac{\lambda_i}{\lambda_{(k)}} \Delta e_i. \tag{20}$$

If Δc_i or Δc_i differs significantly between the retailers within a retailer group, the incentives to allocate units to different transportation options based on their destination (instead of FCFS) increases. In order to evaluate the effects of such allocation, the joint probability mass function of the shipment quantities to all retailers within a retailer group is required. For the single-item case, this pmf is obtained by multinomial disaggregation of M, the total amount of units qualified for shipment in time interval $[t_0-T_k,t_0)$, see Appendix E. The multi-item case can be handled analogously. The joint pmf of the shipment quantities (in Appendix E), also enables evaluation of other more complex transportation costs and emissions structures.

7.   Summary and concluding remarks

This paper studies a distribution system, where replenishments from the central warehouse are consolidated to groups of retailers and dispatched periodically. Shipments are made either using a primary intermodal transportation option where capacity is reserved in advance, or an alternative option where capacity is available on demand. The latter is used only when the reserved capacity is insufficient to transport the entire shipment. Information is centralized and available in real time at the central warehouse. The warehouse replenishes its stock using a continuous review (R_0,Q_0) policy. For this system, we derive the exact probability mass functions for the shipment quantities to each retailer

group. With these distributions at hand, quantity-dependent transportation cost structures and emissions from transportations can be evaluated. Combined with the inventory holding and backorder costs at the warehouse and retailers a joint optimization of inventory replenishment and transportation decisions can be performed with respect to both costs and emissions. A numerical example illustrates how the model can be used for evaluating the cost impact of managerial decisions to impose emissions targets. Single-item as well as multi-item settings can be analyzed and optimized.

Interesting directions for future research include generalizing the model to more flexible demand structures (e.g. compound Poisson), and development of faster heuristics for larger systems. Evaluation of other consolidation policies also offers interesting research venues.

References

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

Axsäter, S. 1993a. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* **41** 777-785.

Axsäter, S. 1993b. Optimization of order-up-to-S policies in two-echelon inventory systems with periodic review. *Naval Research Logistics* **40** 245-253.

Axsäter, S. 1996. Using the deterministic EOQ formula in stochastic inventory control. *Management Science* **42** 830–834.

Axsäter, S. 1997. Simple evaluation of echelon stock (R, Q)-policies for two-level inventory systems. *IIE Transactions* **29** 661–669.

Axsäter, S. 1998. Evaluation of installation stock based (R,Q) policies for two-level inventory systems with Poisson demand. *Operations Research* **46** S135-S145.

Axsäter, S. 2000. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand. *Operations Research* **48** 686-696.

Axsäter, S. 2001. A note on stock replenishment and shipment scheduling for vendor-managed inventory systems. *Management Science* **47** 1306-1310.

Axsäter S., J. Marklund, E.A. Silver. 2002. Heuristic methods for centralized control of one-warehouse, N-retailer inventory systems. *Manufacturing & Service Operations Management* **4** 75-97.

Axsäter, S., J. Marklund. 2008. Optimal "position based" warehouse ordering in divergent two-echelon inventory systems. *Operations Research* **56** 976-991.

Cachon G.P., M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science* **46** 1032-1048.

Çetinkaya, S., J.H. Bookbinder. 2003. Stochastic models for the dispatch of consolidated shipments. *Transportation Research Part B: Methodological* **37** 747-768.

Çetinkaya, S., C.-Y. Lee. 2000. Stock replenishment and shipment scheduling for Vendor-Managed Inventory systems. *Management Science* **46** 217-232.

Çetinkaya, S., E. Tekin, C.-Y. Lee. 2008. A stochastic model for joint inventory and outbound shipment decisions. *IIE Transactions* **40** 324–340.

Chen, F., Y.S. Zheng. 1997. One warehouse multi-retailer systems with centralized stock information. *Operations Research* **45** 275-287.

Chen, F., T. Wang, T.Z. Xu. 2005. Integrated inventory replenishment and temporal shipment consolidation: a comparison of quantity-based and time-based models. *Annals of O. R.* **135** 197-210.

Chu L.Y., Z.M. Shen. 2010. A power-of-two ordering policy for one-warehouse multi-retailer systems with stochastic demand. *Operations Research* **58** 492–502.

Corbett C.J., R.D. Klassen. 2006. Extending the horizons: Environmental excellence as key to improving operations. *Manufacturing and Service Operations Management* **8** 5–22.

Dekker, R., J. Bloemhof, I. Mallidis. 2012. Operations research for green logistics–An overview of aspects, issues, contributions and challenges. *European Journal of Operational Research.* **219** 671-679.

Federgruen, A. 1993. Centralized planning models for multi-echelon inventory systems under uncertainty, in *S.C.Graves et al. Eds., Handbooks in OR & MS* **4**, North Holland Amsterdam, 133-173.

Feller, W. 1968. *An Introduction to Probability Theory and its Applications, Vol. 1*, 3rd edition, John Wiley & Sons, New York.

Forsberg, R. 1995. Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research* **81** 143-153.

Forsberg, R. 1997. Exact evaluation of (R,Q)-policies for two-level inventory systems with Poisson demand. *European Journal of Operational Research* **96** 130-138.

Gaur, V., M. L. Fisher. 2004. A periodic inventory routing problem at a supermarket chain. *Operations Research* **32** 1019–1037.

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

Graves, S.C. 1996. A multi-echelon inventory model with fixed replenishment intervals. *Management Science* **42** 1-18.

Gürbüz, M., K. Moinzadeh, Y.P. Zhou. 2007. Coordinated replenishment strategies in inventory/distribution systems. *Management Science* **53** 293–307.

Houtum G.J. van, K. Inderfurth, W.H.M. Zijm. 1996. Materials coordination in stochastic multi-echelon systems. *European Journal of Operational Research* **95** 1–23.

Howard, C. 2013. New allocation policies for divergent inventory systems with real-time information and shipment consolidation. Working paper Lund University.

Howard, C., J. Marklund. 2011. Evaluation of stock allocation policies in a divergent inventory system with shipment consolidation. *European Journal of Operational Research* **211** 298-309.

Howard, C., O. Stenius. 2013. Partial or complete deliveries in two-echelon inventory systems. Working Paper Lund University.

Kaya, O., D. Kubali, L. Örmeci. 2012. Stochastic models for the coordinated production and shipment problem in a supply chain. *Computers & Industrial Engineering* **64** 838–849.

Kleindorfer, P.R., K. Singhal, L.N. Van Wassenhove, 2005. Sustainable operations management. *Production and Operations Management* **14** 482–492.

Marklund, J. 2002. Centralized inventory control in a two-level distribution system with Poisson demand. *Naval Research Logistics* **49** 798-822.

Marklund, J. 2006. Controlling inventories in divergent supply chains with advance-order information. *Operations Research* **54** 988-1010.

Marklund, J. 2011. Inventory control in divergent supply chains with time-based dispatching and shipment consolidation. *Naval Research Logistics* **58** 59-71.

Marklund, J., Berling, P., 2015, Green Inventory Management, *in Bouchery, Y., Tan, T., Fransoo, J., and Corbett, C. Eds., Sustainable Supply Chains*, Springer, forthcoming.

Marklund, J., K. Rosling. 2012. Lower bounds and heuristics for supply chain stock allocation. *Operations Research* **60** 92–105.

Moinzadeh K. 2002. A multi-echelon inventory system with information exchange. *Management Science* **48** 414–426.

Mutlu, F., S. Çetinkaya, J.H. Bookbinder, 2010. An analytical model for computing the optimal time-and-quantity based policy for consolidated shipments. *IIE Transactions* **42** 367–377.

NTM. 2011. http://www.ntmcalc.org, accessed 08/29/2013.

Özer, Ö. 2003. Replenishment strategies for distribution systems under advanced demand information. *Management Science* **49** 255–272.

Shang, K., Tao, Z., S. Zhou. 2014. Optimizing reorder intervals for two-echelon distribution systems with stochastic demand. *Operations Research*, forthcoming.

Srivastava, S.K. 2007. Green supply-chain management: a state-of-the-art literature review. *International Journal of Management Reviews* **9** 53–80.

Simon, R.M. 1971. Stationary properties of a two echelon inventory model for low demand items. *Operations Research* **19** 761–777.

Stenius, O., G. Karaarslan, J. Marklund, A.G. de Kok. 2015. Exact analysis of divergent inventory systems with time-based shipment consolidation and compound Poisson demand. Working paper Lund University and Eindhoven University of Technology.

Tempelmeier, H., O. Bantel. 2015. Integrated optimization of safety stock and transportation capacity. *European Journal of Operational Research*, forthcoming.

Toth, P., and D., Vigos, *The Vehicle Routing Problem*, SIAM, Philadelphia, 2001.

Zheng, Y.S. 1992. On properties of stochastic inventory systems. *Management Science* **38** 87–103.

Appendix A - Proofs of propositions and lemmas

Proof of Proposition 1. Because the retailer order processes are independent Poisson processes, and the FCFS allocation at the central warehouse, the probability that an arbitrary unit being qualified for shipment in any time interval is destined for a retailer in retailer group k is $\lambda_{(k)}/\lambda_0$. Moreover, this probability is independent of the destination of all other units becoming qualified for shipment. Consequently, the probability of shipment quantity $m_{(k)}$ to retailer group k given m qualified units in total, will be $Bin(m_{(k)}, m, \lambda_{(k)}/\lambda_0)$. Taking the expectation of M renders (2).

Proof of Lemma 1. Because all demanded units eventually will be qualified for shipment, the number of units that will be qualified in time interval $[t_1, t_0)$, M, must be $IL_0^-(t_1)$ plus the demand during $[t_1, t_0)$, $D_0(t_1, t_0)$, minus the amount of demanded units that are not yet qualified for shipment at the end of the interval, $IL_0^-(t_0)$.

Proof of Lemma 2. Every time a unit is demanded at the warehouse the inventory position either decreases by 1 unit or increases by exactly $Q_0 - 1$ units (whenever the inventory position reaches R_0). This means that if the initial inventory position at time t_1 is $IP_0(t_1)$, and the demand in the following time interval $[t_1, t)$ is $D_0(t_1, t)$, the inventory position at the end of this interval must be $IP_0(t_1) - D_0(t_1, t) + n \cdot Q_0$, where n is a non-negative integer. (4) follows as the inventory position always belongs to the interval $[R_0 + 1, R_0 + Q_0]$.

Proof of Proposition 2. The delivery process to the central warehouse is independent of when shipments are made to retailer group k, and all unsatisfied demand is backordered. Thus, the expected number of units that become qualified for shipment to retailer group k per time unit corresponds to the expected demand per time unit $\lambda_{(k)}$. Just after a shipment is made to retailer group k at time t, the number of units qualified for shipment at the central warehouse is 0. In expectation this increases linearly to $T_k\lambda_{(k)}$ units just before the next shipment at $t+T_k$. It follows that the expected number of

units qualified for shipment to retailer group k at the warehouse is $T_k\lambda_{(k)}/2$ and the holding cost for these units is $T_k h_0 \lambda_{(k)}/2$. As all other cost components are larger than or equal to zero, the expected total cost must be larger than $\overline{TC}$ if $T_k > 2\overline{TC}/(\lambda_{(k)}h_0)$.

Appendix B - Determination of a minimum emissions solution

In order to determine TE^e, the minimum expected emissions per time unit, we define $TE_k^e(r)$ to be the lowest possible expected emissions per time unit for retailer group k, given $R_0 = r$. First note that the emissions in retailer group k are independent of the values of **S**, of the shipment intervals and of the capacity reservations of the other groups (T_κ and $w_\kappa \ \forall \ \kappa \neq k$). Hence, TE^e is obtained by solving

$$TE^e = \min_{r=R_0^1,\ldots,R_o^u} \sum_{k=1}^{K} TE_k^e(r). \tag{B1}$$

The lowest emission solution in retailer group k given $R_0=r$, $TE_k^e(r)$, is found by searching through T_k, increasing it incrementally from T_k^1. For each value of T_k and $R_0=r$, TE_k in (16) is evaluated for all capacity reservations, $w_k \in W_k$, constantly updating the smallest value, $\overline{TE}_k$, (when found). Using Lemma B1 below, we stop increasing T_k when $\lambda_{(k)}T_k$ is significantly larger than $\max(w_k)$ (the probability that the shipment quantity is smaller than w_k approaches zero) and TE_k gets sufficiently close to $\lambda_{(k)}\left(\Delta e_{(k)} + \beta_k''/A_k\right)$ for all $w_k \in W_k$. After the search is completed $TE_k^e(r) = \overline{TE}_k$.

Note that Lemma B1 does not state whether TE_k approaches $\lambda_{(k)}\left(\Delta e_{(k)} + \beta_k''/A_k\right)$ from above or below. It is possible with some effort to show that both situations may occur.

Lemma B1.

$$\lim_{T_k\to\infty} TE_k = \lambda_{(k)}\left(\Delta e_{(k)} + \beta_k''/A_k\right). \tag{B2}$$

Proof. Whenever $(m_{(k)} - w_k)^+$ is not a multiple of A_k (where $m_{(k)}$ is the shipment quantity), the load carriers or vehicles on the alternative transportation modes will not be filled completely. Let $Y = n_k A_k - (m_{(k)} - w_k)^+$ be the unused capacity (in units) for the alternative transportation option (where n_k is defined as in (14)). Thus,

$$
\begin{aligned}
\lim_{T_k\to\infty} TE_k &= \lim_{T_k\to\infty} \frac{1}{T_k}\left(\beta_k'(w_k) + \sum_{m_{(k)}=0}^{\infty} P\left(M_{(k)} = m_{(k)}\right)\left[\left(m_{(k)} - w_k\right)^+ \Delta e_{(k)} + n_k \beta_k''\right]\right) \\
&= \lim_{T_k\to\infty} \frac{1}{T_k}\sum_{m_{(k)}=0}^{\infty} P\left(M_{(k)} = m_{(k)}\right)\left[\left(m_{(k)} - w_k\right)\Delta e_{(k)} + \beta_k''\left(m_{(k)} - w_k\right)/A_k + \beta_k'' Y/A_k\right] \\
&= \lim_{T_k\to\infty} \frac{1}{T_k}\left(E\left[m_{(k)} - w_k\right]\left(\Delta e_{(k)} + \beta_k''/A_k\right) + E[Y]\beta_k''/A_k\right) \\
&= \lim_{T_k\to\infty} \frac{1}{T_k}\left(\lambda_{(k)}T_k - w_k\right)\left(\Delta e_{(k)} + \beta_k''/A_k\right) = \lambda_{(k)}\left(\Delta e_{(k)} + \beta_k''/A_k\right).
\end{aligned}
\tag{B3}
$$

The first equality follows from (15) and (16). The second equality is a result of $\beta_k'(w_k)$ being finite for $w_k \in W_k$, and the fact that the probability of a shipment quantity smaller than w_k approaches zero as T_k approaches infinity. The fourth and fifth equalities follow as both $E[Y]\beta_k''/A_k$ and $-w_k\left(\Delta e_{(k)} + \beta_k''/A_k\right)$ are finite for $w_k \in W_k$. ∎

Appendix C – Multi-item systems

Consider J items that are distributed from the central warehouse to the N retailers. The demand of item j at retailer i follows a Poisson process with intensity λ_i^j that is independent of the demand at the other retailers and of the other items. Note that some items might not be available at all retailers, in which case $\lambda_i^j = 0$. Each item j is controlled analogously to the single-item case with base-stock policies at the retailers and a (R,Q) policy at the central warehouse. Again there are $K \leq N$ retailer groups, and shipments are consolidated periodically among all items and all retailers within a retailer group. In addition to the notation for the single-item model we define (t_0 still denotes a time instance when a shipment leaves the warehouse):

Q_0^j = order quantity of item j at the central warehouse

R_0^j = reorder point of item j at the central warehouse, $\mathbf{R}_0 = \{R_0^1, R_0^2,\ldots, R_0^J\}$

S_i^j = order-up-to level of item j at retailer i, $\mathbf{S}^j = \{S_1^j, S_2^j,\ldots, S_N^j\}$, $\tilde{\mathbf{S}} = \{\mathbf{S}^1; \mathbf{S}^2;\ldots; \mathbf{S}^J\}$

L_0^j = replenishment lead-time of item j to the central warehouse

y^j = size (expressed in e.g. weight or volume) of one unit of item j

W_k = set of possible capacity reservation sizes per shipment to retailer group k (expressed in e.g. weight or volume) for the primary transportation option,

$\mathbf{w}$ = vector of capacity reservation per shipment to each of the K retailer groups for the primary transportation option $\{w_1,w_2,\ldots,w_K\}$, $w_k \in W_k \ \forall \ k=1,2,\ldots,K$

h_i^j = holding cost per unit and time unit of item j at stock point i, $i = 0,1,\ldots,N$

b_i^j = backorder cost per unit and time unit of item j at retailer i, $i = 1,2,\ldots,N$

Δc_i = cost increase per size unit (expressed in e.g. weight or volume) for shipping with the alternative transportation mode to retailer i

Δe_i = increase in emissions per size unit (expressed in weight or volume) for shipping with the alternative transportation mode to retailer i

$M_{(k)}^j(t_0)$= shipment quantity (number of units) of item j to all retailers in retailer group k at t_0, $M_{(k)}^j \equiv M_{(k)}^j(t_0)$

$V_{(k)}^j(t_0)=$ shipment size of item j (expressed in weight or volume) to all retailers in retailer group k at t_0 (includes all units of item j shipped), $V_{(k)}^j \equiv V_{(k)}^j(t_0) = y_j M_{(k)}^j$

$V_{(k)}(t_0)=$ shipment size to all retailers in retailer group k at t_0 (includes all units of all items shipped),

$$V_{(k)} \equiv V_{(k)}(t_0) = \sum_{j=1}^{J} V_{(k)}^j$$

$TC(\mathbf{R}_0, \tilde{\mathbf{S}}, \mathbf{T}, \mathbf{w})=$ expected total cost per time unit for the system

$TIC^j(R_0, \mathbf{S}^j, \mathbf{T}, \mathbf{w})=$ expected holding and backorder cost of item j per time unit in the system

$TSC_k(\mathbf{R}_0, \mathbf{T}, \mathbf{w})$ = expected shipment costs per time unit for shipments to retailer group k

$TE(\mathbf{R}_0, \mathbf{T}, \mathbf{w})$ = expected total emissions per time unit for the system

The inventory costs for each item are independent of the reorder points and order-up-to levels of the other items, and can be evaluated with the same methodology as in the single-item case (see Marklund 2011). The total cost function is obtained as

$$TC\left(\mathbf{R}_0, \tilde{\mathbf{S}}, \mathbf{T}, \mathbf{w}\right) = \sum_{j=1}^{J} TIC^j\left(R_0, \mathbf{S}^j, \mathbf{T}, \mathbf{w}\right) + \sum_{k=1}^{K} TSC_k\left(\mathbf{R}_0, \mathbf{T}, \mathbf{w}\right) \tag{C1}$$

For reasons of exposition, we assume that exactly w_k size units can be shipped with the primary transportation mode to retailer group k whenever $V_{(k)} \geq w_k$. (Relaxing this assumption suggests the need for an allocation method that considers how the primary transportation mode is best utilized. Analyzing optimal allocation decisions in this setting is an interesting direction for future research.) As in Section 2, we also assume that $\Delta c_{(k)} = \Delta c_i$ and $\Delta e_{(k)} = \Delta e_i \ \forall i \in \Omega_k$.

The probability mass function of the shipment quantities, $M_{(k)}^j$, can be analyzed independently for all units j=1,2,…,J following the approach in Section 3.1. The probability mass function of the shipment size of all items, $V_{(k)}$, is then obtained by a J-fold convolution of the shipment sizes for the individual items. Clearly, from a numerical perspective, this convolution is an added complication compared to the single-item model.

The expected shipment cost per time unit for shipments to retailer group k can be obtained as

$$TSC_k = \frac{1}{T_k} \sum_{v_{(k)}=0}^{\infty} P\left(V_{(k)} = v_{(k)}\right) SC_k\left(w_k, v_{(k)}\right), \tag{C2}$$

where $SC_k(w_k, v_{(k)})$ is the shipment cost to retailer group k for a realized shipment of size $v_{(k)}$, when w_k size units are reserved on the primary transportation mode. $SC_k(w_k, v_{(k)})$ is obtained as

$$SC_k\left(w_k, v_{(k)}\right) = \alpha_k'\left(w_k\right) + n_k \alpha_k'' + \left(v_{(k)} - w_k\right)^+ \Delta c_{(k)}, \tag{C3}$$

where n_k is the integer that satisfies $(n_k - 1)A_k < \left(v_{(k)} - w_k\right)^+ \leq n_k A_k$. The emissions of a shipment with size $v_{(k)}$ to retailer group k, when w_k size units are reserved on the primary transportation mode, $E_k(w_k, v_{(k)})$, is obtained as

$$E_k\left(w_k, v_{(k)}\right) = \beta_k'\left(w_k\right) + n_k \beta_k'' + \left(v_{(k)} - w_k\right)^+ \Delta e_{(k)}. \tag{C4}$$

This renders the expected emissions per time unit for shipments to retailer group k

$$TE_k = \frac{1}{T_k} \sum_{v_{(k)}=0}^{\infty} P\left(V_{(k)} = v_{(k)}\right) E_k\left(w_k, v_{(k)}\right), \tag{C5}$$

and the total expected emissions per time unit

$$TE\left(\mathbf{R}_0, \mathbf{T}, \mathbf{w}\right) = \sum_{k=1}^{K} TE_k. \tag{C6}$$

The optimization can be performed analogously to the single-item case. For each relevant combination of $\mathbf{T}$, $\mathbf{w}$ and $\mathbf{R}_0$, check if the solution is environmentally feasible with respect to the emissions constraint. If $TE(\mathbf{R}_0, \mathbf{T}, \mathbf{w}) < \theta$, find the optimal order-up-to level for each item j and each retailer i, S_i^j, using that the total cost is separable and convex in the retailer order up to levels $\mathbf{\tilde{S}} = \{S_1^1, S_2^1, \ldots, S_N^1; S_1^2, S_2^2, \ldots, S_N^2; \ldots; S_1^J, S_2^J, \ldots, S_N^J\}$. If a new lowest cost is found, $\overline{TC}$ is updated. When the search space (defined by bounds on $\mathbf{T}$ and $\mathbf{R}_0$, and the finite set of possible capacity reservations, $\mathbf{W_k}$ for all k) is exhausted, the optimal solution has been found and the minimum total cost is $\overline{TC}$.

The bounds on T_k and R_0 presented in Section 4 also hold for the multi-item case. If there is a need to find the lowest possible expected emissions per time unit, TE^e, we define $\mathbf{R}^J$ as the set of combinations of $\mathbf{R}_0$, where the warehouse reorder point for item j, R_0^j, is between the lower and upper bound for all $j = 1, \ldots J$. We also define $TE_k^e(\mathbf{r})$ as the minimum expected emissions in retailer group k for a given $\mathbf{R}_0 = \mathbf{r}$. $TE_k^e(\mathbf{r})$ can be found analogously to $TE_k^e(r)$ in the single-item case. TE^e is then

$$TE^e = \min_{\mathbf{r} \in \mathbf{R}^J} \sum_{k=1}^{K} TE_k^e\left(\mathbf{r}\right), \tag{C7}$$

Appendix D – Different transportation lead-times

This Section analyzes situations where the transportation lead-time to retailer i with the primary transportation option, L_i^P, differs from the transportation lead-time of the alternative transportation option, L_i^A (for some or all of the retailers). This means that a shipment where both transportation options are used to transport units to retailer i is delivered at two separate time instances. We assume that order crossing is not allowed, which means that a shipment dispatched to retailer i at time t will be delivered before any shipments dispatched to this retailer after time t. To ensure this, the shipment interval, T_k, is required to be at least as long as the difference between the transportation lead-times L_i^P and L_i^A for all retailers, $T_k \geq \max_{i \in \Omega_k} \left| L_i^P - L_i^A \right|$. Consistent with the FCFS policy in use, the units in a shipment that were demanded first will be transported on the faster transportation option.

We focus on situations where the transportation lead-time of the primary option is longer than for the alternative option. Arguably this is more plausible than the opposite when the primary option is an intermodal train solution, and the alternative is direct truck transports. The opposite situation can be analyzed analogously. Also, the analysis focuses on the single-item model, but extending it to multi-item systems is straightforward. The presented solution is based on first determining the costs for the *base model* (analyzed so far) where the transportation lead-times for the alternative option is the same as for the primary option across all retailers. The second step is to determine the expected cost increase per time unit associated with having a faster alternative transportation option. Given that the other system parameters and decision variables are the same, this change does not affect the shipment quantity (and thereby the transportation costs and emissions) or the holding cost at the central warehouse. The changes will be isolated to the holding and backorder costs at the retailers and the holding costs during transportation. The latter is excluded from the *base model* as it is there a constant term, unaffected by the decision variables. Let,

$\Delta C_i(R_0, \mathbf{S}, \mathbf{T}, \mathbf{w})$ = Expected cost increase per time unit for a system where the units transported with the alternative option to retailer i have a transportation lead-time L_i^A $\forall i \in \Omega$, compared to the *base model* where all units have a transportation lead-time L_i^P $\forall i \in \Omega$ ($\Delta C_i < 0$ is equivalent to an expected cost saving).

The total expected cost per time unit of the system is then

$$TC(R_0, \mathbf{S}, \mathbf{T}, \mathbf{w}) = TIC(R_0, \mathbf{S}, \mathbf{T}) + \sum_{i \in \Omega} \Delta C_i(R_0, \mathbf{S}, \mathbf{T}, \mathbf{w}) + TSC(R_0, \mathbf{S}, \mathbf{T}). \tag{D1}$$

The same optimization procedure as before can be used. The retailer costs are still be separable and convex in the retailer order-up-to levels (the holding costs during transports are unaffected by the retailer order-up-to levels). What remains to be analyzed is how the holding costs during transportation and the holding and backorder costs at retailer i, for a given set of system parameters and decision variables, are affected by decreasing the transportation lead-time of the alternative transportation option from L_i^P to L_i^A, i.e. to obtain $\Delta C_i(R_0, \mathbf{S}, \mathbf{T}, \mathbf{w})$. Focusing on time t_0, when an arbitrary shipment is dispatched from the warehouse to retailer i, we define:

M_i^P =   Number of units shipped to retailer i by the primary transportation option at t_0

M_i^A =   Number of units shipped to retailer i by the alternative transportation option at t_0

B_i =   Backorders at the central warehouse at t_0 designated to retailer i

Φ_i = (M_i^P, M_i^A, B_i) = *Shipment state* associated with retailer i at t_0

$\Delta RC_i(\Phi_i)$ =   Expected increase of holding and backorder costs for the units dispatched to retailer i at t_0 compared to the *base model*, for *shipment state* Φ_i.

Δh_i = decrease in holding cost during transport for a unit transported with the alternative option instead of primary option to retailer i. May, for example, be determined as $(L_i^P - L_i^A)$ multiplied with some appropriate holding cost rate per unit and time unit.

Ψ_i = *Customer demand number* at retailer i. When $\Psi_i > 0$, it indicates the Ψ_i^{th} customer demand occurring after (or exactly at) time t_0 at retailer i,, and when $\Psi_i \leq 0$, it indicates the $(-\Psi_i +1)^{th}$ customer demand occurring prior to time t_0 at retailer i.

$\Delta HC_i(\Psi_i)$ = Expected increase in holding cost at retailer i for a unit satisfying the demand with *customer demand number* Ψ_i, when the transportation lead-time is L_i^A instead of L_i^P

$\Delta BC_i(\Psi_i)$ = Expected decrease in backorder costs at retailer i for a unit satisfying the demand with *customer demand number* Ψ_i, when the transportation lead-time is L_i^A instead of L_i^P

$G_i^\Psi(t)$ = Erlang(ψ,λ_i) cumulative distribution function, $\psi \geq 0$, $G_i^0(0) = 1$

$g_i^\psi(t)$ = Erlang(ψ,λ_i) probability density function,

The expression for the expected cost increase per time unit of a shorter transportation lead-time for the alternative option at retailer i, $\Delta C_i(R_0,\mathbf{S},\mathbf{T},\mathbf{w})$, in (D2) is based on analyzing the *shipment state* to retailer i at t_0, $\Phi_i = (M_i^P,M_i^A,B_i)$. For a given *shipment state* ($M_i^P = m_i^P$, $M_i^A = m_i^A$ and $B_i = x_i^B$), we can evaluate the expected holding and backorder cost increase for the units transported with the alternative option compared to the *base model*, $\Delta RC_i(m_i^P,m_i^A,x_i^B)$. Expectation over the relevant set of shipment states renders the expected cost difference of an arbitrary shipment. The expected cost increase per time unit is then obtained after division with the shipment interval, T_k.

$$\Delta C_i\left(R_0,\mathbf{S},\mathbf{T},\mathbf{w}\right) = \frac{1}{T_k}\sum_{m_i^P=0}^{w_i}\sum_{m_i^A=1}^{\infty}\sum_{x_i^B=0}^{\infty}P\left\{\Phi_i = (m_i^P,m_i^A,x_i^B)\right\}\Delta RC_i\left(m_i^P,m_i^A,x_i^B\right) \tag{D2}$$

Note, because $\Delta RC_i(m_i^P,0,x_i^B) = 0 \;\forall m_i^P,x^B$ we only need to consider the cases where $m_i^A > 0$. How to determine $\Delta RC_i(\Phi_i)$ is described in Section D.1, while the *shipment state* probability, $P\left\{\Phi_i = (m_i^P,m_i^A,x_i^B)\right\}$, for all relevant outcomes is determined in Section D.2.

D.1. Analysis of $\Delta RC_i(\Phi_i)$

Lemma D1. The units shipped with the alternative transportation option to retailer i at t_0 will satisfy customer demands with *customer demand numbers*, $\Psi_i \in [S_i - B_i - M_i^P - M_i^A + 1,\; S_i - B_i - M_i^P]$.

Proof. Let $\tilde{M}_i$ be the amount of units in transit to retailer i at t_0 excluding the shipment at t_0. By definition, the inventory position at retailer i at t_0, S_i, is the inventory level, $IL_i(t_0)$, plus all outstanding orders. The outstanding orders just after the shipment at t_0 can be expressed as $B_i + M_i^P + M_i^A + \tilde{M}_i$. Hence, $IL_i(t_0) + \tilde{M}_i = S_i - B_i - M_i^P - M_i^A$. Due to the FCFS allocations, if $IL_i(t_0) + \tilde{M}_i \geq 0$, the first

27

unit from the shipment at t_0 that arrives to retailer i will satisfy the $(IL_i(t_0) + \tilde{M}_i + 1)^{th}$ demand at retailer i occurring after (or exactly at) t_0. Similarly, if $IL_i(t_0) + \tilde{M}_i < 0$, the first unit from the shipment at t_0 that arrives to retailer i will satisfy the $(IL_i(t_0) + \tilde{M}_i)^{th}$ demand at retailer i prior to t_0. Thus, the units shipped at t_0 will satisfy the customer demands at retailer i with *customer demand numbers*, Ψ_i, between $S_i - B_i - M_i^P - M_i^A + 1$ and $S_i - B_i$. Invoking the assumption (made in accordance with the FCFS principle) that the units demanded first are shipped with the faster alternative transportation option and are thereby delivered first, completes the proof. ■

Based on Lemma D1, $\Delta RC_i(\Phi_i)$, can be determined as

$$\Delta RC_i\left(\Phi_i\right) = -M_i^A \Delta h_i + \sum_{\psi = S_i - B_i - M_i^P - M_i^A + 1}^{S_i - B_i - M_i^P} \Delta HC_i\left(\psi\right) - \Delta BC_i\left(\psi\right) \tag{D3}$$

When analyzing the per unit holding cost increase, $\Delta HC_i(\Psi_i)$, compared to the *base model*, we first consider situations where $\Psi_i > 0$, i.e. the customer demand occurs after (or exactly at) time t_0, see (D4). Let τ denote the time until the Ψ_i^{th} unit after (or exactly at) time t_0 is demanded, and note that it is Erlang(Ψ_i, λ_i)-distributed. Figure D1 illustrates the events on a time line and the associated cost differences for a unit (given $L_i^A < \tau < L_i^P$ for $\Psi_i > 0$).
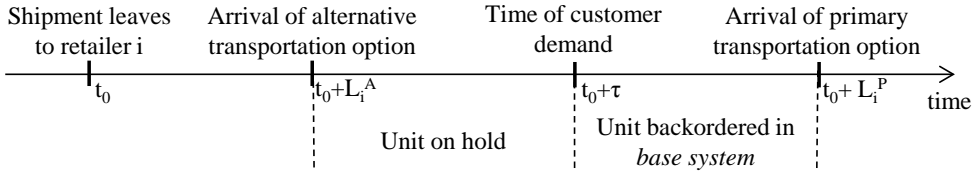
Figure D1. Illustration of inventory cost differences incurred on a unit when $L_i^A < \tau < L_i^P$.

For $\tau \leq L_i^A$ the unit will be demanded by a customer before it arrives to retailer i, regardless if the transportation lead-time is L_i^P or L_i^A, and there is consequently no holding cost increase. If $L_i^A < \tau < L_i^P$ (see Figure D1) the unit in the *base model* will be demanded before it arrives at retailer i and will therefore not be on stock at retailer i. However, with transportation lead-time L_i^A the unit will incur a holding cost for $\tau - L_i^A$ time units at retailer i. When $\tau \geq L_i^P$ the unit will be on stock at retailer i for $L_i^P - L_i^A$ time units longer if the transportation lead-time is L_i^A instead of L_i^P. We get

$$\Delta HC_i\left(\Psi_i\right) = \int_{L_i^A}^{L_i^P} g_i^{\tau}\left(\Psi_i\right)\left(\tau - L_i^A\right) h_i d\tau + \left(1 - G_i^{L_i^P}\left(\Psi_i\right)\right)\left(L_i^P - L_i^A\right) h_i \qquad , \Psi_i > 0 \tag{D4}$$

For $\Psi_i \leq 0$, the demand has occurred before or exactly at time t_0 and there are no holding costs incurred on the units in neither cases and therefore

$$\Delta HC_i\left(\Psi_i\right) = 0 \qquad , \Psi_i \leq 0 \tag{D5}$$

Analogously for the backorder cost decrease, if $\tau \leq L_i^A$ the unit will be backordered for $L_i^P - L_i^A$ time units longer in the *base model* than if the transportation lead-time is L_i^A. If $L_i^A < \tau < L_i^P$, the unit will not be backordered when the transportation lead-time is L_i^A (see Figure D1), but for the *base model* it will be backordered for $L_i^P - \tau$ time units. When $\tau \geq L_i^P$ the unit will never be backordered.

$$\Delta BC_i(\Psi_i) = G_i^{L_i^A}(\Psi_i)(L_i^P - L_i^A)b_i + \int_{L_i^A}^{L_i^P} g_i^\tau(\Psi_i)(L_i^P - \tau_i)b_i d\tau \qquad , \Psi_i > 0 \qquad (D6)$$

For $\Psi_i \leq 0$, the demand has occurred before or exactly at time t_0 and the units are backordered already at t_0. We thus get

$$\Delta BC_i(\Psi_i) = (L_i^P - L_i^A)b_i \qquad , \Psi_i \leq 0 . \qquad (D7)$$

D.2. Probability of the shipment state $P\{\Phi_i = (m_i^P, m_i^A, x_i^B)\}$

The probabilities of the shipment states are determined in a three-step procedure. The first step is to determine the joint probability of m qualified units to all retailers (in all retailer groups) between t_1 and t_0, and of y backorders at the central warehouse at t_0, $P\{M = m \text{ and } IL_0^-(t_0) = y\}$. This can be done using the same technique presented for obtaining $P\{M = m\}$ in Section 3.1. For $L_0 \leq T_k$, $P\{M = m \text{ and } IL_0^-(t_0) = y\}$ is obtained using (6), (7) and (8), with the technique described after (8) in Section 3.1. Note that for each value of x, D_A, D_B and D_C, the value of both M and $IL_0^-(t_0)$ are uniquely defined. For $L_0 > T_k$, $P\{M = m \text{ and } IL_0^-(t_0) = y\}$ is obtained using the same technique, but with (10) and (11) determining the values of $IL_0(t_0)$ and M.

In step 2, $P\{M_{(k)} = m_{(k)} \text{ and } IL_0^-(t_0) = y\}$ is determined. Following the same logic as in Proposition 1 and because the amount of backorders at the central warehouse at t_0 is independent of which retailer a customer arrives to, this probability can be obtained by binomial disaggregation. Thus,

$$P\{M_{(k)} = m_{(k)} \text{ and } IL_0^-(t_0) = y\} = \sum_{m=m_{(k)}}^{\infty} P\{M = m \text{ and } IL_0^-(t_0) = y\} Bin(m_{(k)}, m, \lambda_{(k)}/\lambda_0). \qquad (D8)$$

In the final step, the joint distribution of $M_i^P = m_i^P$, $M_i^A = m_i^A$ and $B_i = x_i^B$ is determined. Because we are only interested in the cases where there is at least one unit shipped with the alternative transportation option, we require that $M_{(k)}$ is at least $w_k + 1$. When $M_{(k)} = m_{(k)}$ ($m_{(k)} > w_k$) and $IL_0^-(t_0) = y$, the primary transportation option to retailer group k is fully utilized and the amount of units on the primary transportation option to retailer i, M_i^P, is binomially distributed

$$P\{M_i^P = m_i^P \mid M_{(k)} = m_{(k)} \text{ and } IL_0^-(t_0) = y\} = Bin(m_i^P, w_k, \lambda_i/\lambda_{(k)}). \qquad (D9)$$

Given that $M_{(k)} = m_{(k)}$ ($m_{(k)} \geq w_k + M_i^A$) and $IL_0^-(t_0) = y$, the amount of units to all retailers in retailer group k shipped on the alternative transportation option is $m_{(k)} - w_k$, and the amount of units transported with the alternative option to retailer i, M_i^A, is binomially distributed

$$P\left\{M_i^A = m_i^A \middle| M_{(k)} = m_{(k)} \text{ and } IL_0^-(t_0) = y\right\} = Bin\left(m_i^A, m_{(k)} - w_k, \lambda_i/\lambda_{(k)}\right). \tag{D10}$$

Moreover, when $M_{(k)} = m_{(k)}$ ($m_{(k)} \geq w_k + M_i^A$) and $IL_0^-(t_0) = y$, the amount of backorders designated to retailer i, B_i, is binomially distributed

$$P\left\{B_i = x_i^B \middle| M_{(k)} = m_{(k)} \text{ and } IL_0^-(t_0) = y\right\} = Bin\left(x_i^B, y, \lambda_i/\lambda_0\right). \tag{D11}$$

Furthermore, given $M_{(k)} = m_{(k)}$ ($m_{(k)} \geq w_k + M_i^A$) and $IL_0^-(t_0) = y$, the distributions of $M_i^P = m_i^P$, $M_i^A = m_i^A$ and $B_i = x_i^B$ are independent rendering the final expression (D12). The independence, and the binomial disaggregation follow from the Poisson demand processes and FCFS allocations.

$$P\left\{M_i^P = m_i^P, M_i^A = m_i^A \text{ and } B_i = x_i^B\right\} =$$

$$\sum_{m_{(k)}=w_k+m_i^A}^{\infty} \sum_{y=x_i^B}^{\infty} \left[\begin{array}{l} P\left\{M_{(k)} = m_{(k)} \text{ and } IL_0^-(t_0) = y\right\} \cdot \\ Bin\left(m_i^P, w_k, \lambda_i/\lambda_{(k)}\right) Bin\left(m_i^A, m_{(k)} - w_k, \lambda_i/\lambda_{(k)}\right) Bin\left(x_i^B, y, \lambda_i/\lambda_0\right) \end{array} \right]. \tag{D12}$$

Appendix E - The joint pmf of shipment quantities

In this Appendix we show how to determine the joint probability mass function of the shipment quantities to all retailers within retailer group k. Knowledge of this probability distribution enables extensions of the model to other (retailer dependent) cost and emissions structures, and consideration of different allocation decisions. The latter arising, for instance, when the cost increase, Δc_i, and the emissions increase, Δe_i, is not the same for all retailers within a retailer group (as assumed in the main analysis), and the shipment quantity exceeds the reserved capacity to retailer group k.

$M_i(t_0) =$ Shipment quantity to retailer i, $i \in \Omega_k$, (i.e. number of units shipped to retailer i in retailer group k at t_0), $M_i \equiv M_i(t_0)$

$\mathbf{M}_k = \{M_1, M_2, \ldots, M_{N_k})$

$\mathbf{m}_k = \{m_1, m_2, \ldots, m_{N_k}\}$,

$\boldsymbol{\lambda}_k = \{\lambda_1, \lambda_2, \ldots, \lambda_{N_k}\}$

Because of the Poisson demand processes, a unit that becomes qualified for shipment in any time interval is destined for retailer i with probability λ_i/λ_0, and in analogy to Proposition 1, the joint probability $P(\mathbf{M}_k = \mathbf{m}_k)$ can be obtained by multinomial disaggregation of M, the total amount of units qualified for shipment in time interval $[t_0-T_k, t_0)$. Thus, $P\{\mathbf{M}_k = \mathbf{m}_k | M=m\}$ follows a multinomial

distribution (see, for example, Feller (1968) for a definition). Letting $\varphi = \sum_{i=1}^{N_k} m_i$, and recalling that

$\lambda_{(k)} = \sum_{i=1}^{N_k} \lambda_i$ denotes the total demand intensity in retailer group k, (E1) and E2) follows

$$P\{\mathbf{M}_k = \mathbf{m}_k \,|\, M = m\} = \frac{m!}{(m-\varphi)! \prod_{i=1}^{N_k} m_i!} \left(\frac{\lambda_0 - \lambda_{(k)}}{\lambda_0}\right)^{(m-\varphi)} \prod_{i=1}^{N_k} \left(\frac{\lambda_i}{\lambda_0}\right)^{m_i}, \qquad (E1)$$

$$P\{\mathbf{M}_k = \mathbf{m}_k\} = \sum_{m=\varphi}^{\infty} P\{M = m\} P\{\mathbf{M}_k = \mathbf{m}_k \,|\, M = m\}, \qquad (E2)$$

Exact Analysis of Divergent Inventory Systems with Time-Based Shipment Consolidation and Compound Poisson Demand

Olof Stenius

Department of Industrial Management and Logistics, Lund University, Box 118 22100 Lund, Sweden, e-mail: Olle.Stenius@iml.lth.se

Ayşe Gönül Karaarslan

Department of Econometrics, Erasmus University Rotterdam, PO Box 1738 3000DR Rotterdam, The Netherlands, Tel: +31-10-4081342, e-mail: Karaarslan@ese.eur.nl

Johan Marklund

Department of Industrial Management and Logistics, Lund University, Box 118 22100 Lund, Sweden, Tel: +46-46-2228013, e-mail:Johan.Marklund@iml.lth.se

A. G. de Kok

Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, P.O. Box 513 5600MB Eindhoven, The Netherlands, Tel: +31-40-2473849, e-mail: A.G.d.Kok@tue.nl

Abstract

Sustainable and efficient management of a distribution system requires coordination between transportation planning and inventory control decisions. In this context, we consider a one warehouse multi-retailer inventory system with a time-based shipment consolidation policy at the warehouse. This means that there are fixed costs associated with each shipment, and retailer orders are consolidated and shipped periodically to groups of retailers sharing the same delivery routes. Customer demand is compound Poisson distributed and unsatisfied demand at each stockpoint is backordered and allocated on a First-Come First-Served basis. The system is centralized and inventory levels are reviewed continuously. The warehouse has access to real-time inventory information from the retailers, and uses a (R, nQ) policy to replenish from an outside supplier/manufacturer. We derive the exact probability distributions for the inventory levels at the retailers, and use these to obtain exact expressions for the system's expected shipment, holding and backorder costs, its average inventory levels and fill rates. Based on the analytical properties of the objective function, we construct an optimization procedure by deriving bounds on the optimal reorder levels and shipment intervals both for single-item and multi-item systems.

Keywords: Inventory, Multi-echelon, Multi-item, Stochastic, Shipment consolidation, Continuous review, Compound Poisson demand

1 Introduction

The technological development of integrated supply chain information systems is making real-time point-of-sale and inventory information more and more accessible, also across multi-tier supply chains. In the process, administrative costs of sharing information and placing of orders are decreasing, thus reducing the economic incentives for aggregating demand information into orders. On the other hand, set up costs, batch processing, and shipment consolidation are important considerations in manufacturing and distribution of physical products. In freight transportation these issues are accentuated by fluctuating fuel prices and increasing emphasis on environmental concerns and sustainability. With respect to the latter, shipment consolidation can reduce the number of shipments (e.g. trucks) and thereby achieve both lower transportation costs and lower (carbon) emissions. However, consolidation typically means longer replenishment lead times and increased inventory costs. Thus, sustainability in terms of economic viability and environmental friendliness requires coordination and balancing of shipment and inventory decisions.

In this paper, we focus on these issues in the context of a divergent two-echelon inventory system with a central warehouse and multiple non-identical retailers facing compound Poisson demand. The system is characterized by a time-based shipment consolidation policy at the warehouse, in conjunction with real-time point of sale data and centralized inventory information. The consolidation policy means that outbound shipments from the central warehouse are dispatched periodically to groups of retailers sharing the same delivery routes (a group may consist of a single retailer). Thus, retailer orders are consolidated for the retailers on the same route over the time between shipments, referred to as the shipment interval. The delivery routes and grouping of retailers are determined exogenously and optimizing their design is beyond the scope of this paper. Inventory replenishments are based on continuous review information, batch ordering at the warehouse, and base stock ordering at the retailers. The contributions of our work include derivations of the probability distributions for all inventory levels in the system, and determination of the associated expected costs and retailer fill rates. We also provide optimality bounds that allow for joint optimization of the shipment intervals and the reorder levels in the system. The results encompass both the case of backorder costs per unit and time unit at the retailers, and the case of fill rate constraints. For the latter, the total costs only consist of the expected shipment and holding costs. The analysis is applicable to single-item as well as multi-item systems, although the focus in this work is placed on the former. The approach for analyzing the system is new, and can be used for analyzing other types of divergent inventory systems.

Initial motivation for our work stems from discussions with a large European company that manufactures and sells sheet metal products to the construction industry. The company produces both to order and to stock, where the former typically involves customized solutions for large construction projects. Focusing on the latter, products are distributed from a central warehouse to a large number of retailers and market companies (hereafter jointly referred to as retailers) using connected IT and inventory control systems. This means that there is centralized access to point-of-sale data and inventory information in real time for the network. Currently, this information is not used for centralized control of the system, but there are ongoing discussions about how this may be done. Many of the company's products are large and bulky with relatively low value to volume ratio. Thus, transportation costs are generally high, meaning that shipment consolidation, and efficient use of the transport capacity are important for the company. Consolidated shipments from the warehouse to the retailers are dispatched periodically to groups of retailers. The transportation planning is performed in-house, but the transports are contracted from external service providers. All outbound transports from the warehouse are made by truck, either a dedicated vehicle delivering to a group of retailers on a specified route, or a shared vehicle utilizing the transport provider's network of terminals. Looking downstream, the majority of the company's customers are construction companies of different sizes, which typically place batch orders dedicated to specific construction projects. As these projects vary in sizes so do the batch sizes. Thus, the customer demand is characterized by randomness both in the number of orders that arrive, and in the size of individual orders. Structurally, this corresponds well to a compound Poisson process, which motivates its use in our present work. Upstream, the warehouse replenishes most of its stock by placing orders with the company's own manufacturing plants.

Further motivation for our work is based on contacts with several other companies (including some spare parts service providers) that use, or are in the position to use, different types of VMI (Vendor Managed Inventory) systems in combination with periodic shipment schedules. (For overviews of different types of VMI systems, see, for example, Cheung and Lee 2002 and Marques et al. 2010.) Although the details may vary, these systems are characterized by a supplier/central warehouse with the mandate to control inventories at different customer locations/retailers under specified service agreements. This control is typically facilitated by access to point-of-sale data and inventory information from the retailers. These centralized systems offer the supplier flexibility in planning production, distribution and replenishing activities. However, a recurring question these companies struggle with is how the centralized inventory information may be used for improved control of their multi-echelon

inventory systems.

Literature Review

Our work is closely related to Marklund (2011), which considers a similar system under the more restrictive assumption of Poisson demand. Marklund (2011) presents a fast recursive procedure for determining the exact average costs per time unit, and for optimizing the reorder levels at all inventory locations for a given set of shipment intervals. A heuristic for determining near optimal shipment intervals is also presented. Compared to Marklund (2011) our present work distinguishes itself in three major ways. Firstly, it is more general as we provide an exact approach for compound Poisson demand. Secondly, we use a new methodology for analyzing the system, which enables determination of the inventory level distributions and service levels at all locations, not just the expected costs as in Marklund (2011). Thirdly, we provide upper and lower bounds for the optimal shipment intervals and reorder points in the system, thereby enabling joint optimization of these decision variables.

Howard and Marklund (2011) and Howard (2013) build on the work in Marklund (2011) and use simulation to investigate the cost benefits of replacing the First-Come First-Served (FCFS) allocation assumption at the warehouse with state dependent myopic allocation policies under Poisson demand. Howard and Marklund (2011) deal with one retailer group whereas Howard (2013) considers multiple retailer groups and a policy that offers a performance guarantee over FCFS. Both studies conclude that some cost benefits exist for long retailer lead times, particularly, when the allocation decision is postponed to the moment of delivery, but in general the FCFS assumption performs very well. Gürbüz et al. (2007) also consider joint inventory and transportation decisions, but for a system where the warehouse is a stockless cross docking facility that orders for a set of retailers simultaneously. This policy is compared to three well-known policies including one with fixed replenishment intervals.

Our present work is also related to the general multi-echelon literature which does not take shipment consolidation decisions into consideration. Of particular interest is the stream of literature on analysis of continuous review one-warehouse multiple-retailer systems. For overviews, see for example, Axsäter (2003) and Axsäter and Marklund (2008). Early contributions include Simon (1971), Graves (1985) and Axsäter (1990), which consider systems with Poisson demand, complete backordering, FCFS allocation, and base-stock policies at all locations. Graves (1985) uses an exact approach similar to Simon (1971) for determining the probability distributions of the retailers inventory levels and the expected system costs. An accurate approximation is also presented. The analysis is based

on determining the steady state probability distributions of the number of outstanding orders at each retailer by binomial disaggregation of the total number of outstanding orders. The approach is also commonly translated into disaggregation of the total number of warehouse backorders to obtain the distribution of warehouse backorders destined for each retailer (c.f., Axsäter 2006). Generalizing the method to handle compound Poisson demand is difficult and has so far not been done. The challenge is how to disaggregate the distribution of the total number of outstanding orders to specific retailers. Axsäter (1990) provides exact and computationally efficient recursive expressions for the system's expected inventory holding and backorder costs. His approach is based on tracking an arbitrary unit through the system, and determining the holding and backorder costs it incurs. The method does not involve the inventory level distributions at the retailers. Our work is related to both these approaches. As we are interested in the inventory levels in the system, we derive the exact probability distributions of warehouse backorders destined to each retailer. However, instead of disaggregating the total amount of backorders, we obtain these distributions by tracking what happens in the system backwards and forwards in time. The methodology of Axsäter (1990) has been extended in various ways to deal with more general divergent systems. For example, Axsäter (1993a, 1998) and Forsberg (1997) consider exact and approximate models with installation stock (R, Q)-policies and Poisson demand. In case of compound Poisson demand, Forsberg (1995) shows how to exactly calculate costs for base-stock policies at the retailers as weighted sums of the expected costs of (S-1,S) systems with Poisson demand obtained from Axsäter (1990). Axsäter (1997) extends the model to echelon stock (R, Q)-policies (the echelon stock of an inventory location includes the stock at the location itself and at all downstream locations). In a parallel work, Chen and Zheng (1997) provide an alternative method for evaluating echelon-stock (R, Q)-policies that is exact for Poisson demand and approximate for compound Poisson demand. This approach is related to the disaggregation of warehouse backorders used by Simon (1971) and Graves (1985). Of particular interest for our work is Axsäter (2000), which provides exact analysis of the expected holding and backorder costs and the probability distributions for the inventory levels in a one-warehouse N-retailer system with installation stock (R, Q)-policies and compound Poisson demand. In the special case of no shipment consolidation and batch quantities equal to one at all retailers, our present work offers a different way to analyze the system in Axsäter (2000).

The literature on divergent continuous review systems also contains papers that investigate other warehouse ordering policies utilizing more detailed inventory and demand information. Marklund (2002) provides an exact analysis of a new type of service level policy at the warehouse (referred

to as an (α_0, Q_0) policy) that uses real-time information about the individual inventory positions at all stockpoints. In a parallel work Moinzadeh (2002) investigates a generalized installation-stock (R, Q) policy at the warehouse assuming identical retailers. Axsäter and Marklund (2008) derive a warehouse ordering policy that is optimal in the class of "position based" policies, which encompass all the policies discussed above. They also relax the FCFS assumption made in all previous models cited above. Marklund (2006) focuses on the use of advance order information, and provides exact and approximate analysis of warehouse reservation policies. These policies enable the warehouse to differentiate its service to the retailers through temporal allocation and prioritization. All the models above assume the use of partial deliveries, i.e., all stockpoints ship what is available as soon as possible.

Apart from the exact results mentioned above there exist a large number of approximations for analyzing one-warehouse-N-retailer inventory systems, (c.f., Axsäter (2003) for an overview). One of the most common approximation techniques is to replace the stochastic delay due to stockouts by its average value. This idea originates from the METRIC model by Sherbrooke (1968) and has been developed and adapted to many different settings in e.g. Sherbrooke (1986), Zipkin (2005 p.335), Andersson et al. (1998), Andersson and Marklund (2000), Axsäter (2003) and Berling and Marklund (2013, 2014). This group of approximations is related to our work as the average waiting time usually is obtained by dividing the average total amount of backorders at the central warehouse with the average total demand per time unit. In some special cases this renders the correct mean, but in general the average waiting time per demanded unit at a warehouse is different across the retailers. Our analysis of the backorders at the central warehouse can be used for determining the correct average waiting time for each retailer. In this respect our work is related to Kiesmüller et al. (2004), which focus on an approximation model involving the first two moments of the waiting time.

The time-based dispatching and use of fixed shipment intervals links our present work to the research on divergent periodic review systems. A major difference compared to the traditional periodic review literature (see, for example, Federgruen and Zipkin 1984, Jackson 1988, Federgruen 1993, Axsäter 1993b, Graves 1996, Houtum et al. 1996, Heijden et al. 1997, Diks and de Kok 1998,1999, Cachon 1999, Cachon and Fisher 2000, Axsäter et al. 2002, Özer 2003, Chu and Shen 2010, Marklund and Rosling 2012, Shang et al. 2014, and references therein) is that in our current model only shipments are made periodically, while inventory is reviewed and replenished continuously. Graves (1996), Axsäter (1993b), and Shang et al. (2014) have a closer relationship with our work because they assume a virtual (FCFS) allocation policy based on Poisson demand arrivals. This means that

the inventory levels must be monitored continuously even though orders are generated periodically. Our model assumptions are less restrictive with respect to the compound Poisson demand and the (R, Q)-policy at the warehouse. A distinguishing feature of Shang et al. (2014) compared to Graves (1996) and Axsäter (1993b), is that not only the base-stock levels but also the reorder intervals are decision variables. An important contribution is that the authors are able to provide bounds that facilitate optimization of all these decision variables to minimize the total expected costs. The reorder intervals are related to the shipment intervals in our model.

Finally, there is a connection between our work and the body of literature that investigates different types of consolidation policies for outbound shipments in a single-echelon context. The focus in this literature is placed on a VMI supplier (typically with a negligible replenishment lead time) that receives orders from multiple retailers. The supplier wants to decide how and when to replenish and dispatch shipments in order to minimize its inventory and shipment costs. Examples from this literature include Çetinkaya and Lee (2000), Axsäter (2001), Çetinkaya and Bookbinder (2003), Çetinkaya et al. (2008), and Mutlu et al. (2010). In principle, these (and other) papers consider three types of consolidation strategies for Poisson or compound Poisson demand: (i) time-based consolidation policies where time dictates when shipments are dispatched, (ii) quantity-based policies where shipments are consolidated into fixed dispatch quantities, and (iii) time-and-quantity based policies where shipments leave either when a dispatch quantity is reached or when a certain time has passed since the last dispatch. None of these papers provide an exact analysis similar to our present work.

The remainder of this paper is organized as follows. Section 2, describes the detailed model assumptions and the associated total cost function for the single-item model. Section 3 presents exact analysis of the inventory level distributions as well as the backorder distributions at the central warehouse. Section 4 explains the optimization procedure. Section 6 extends the model formulation, cost analysis and optimization procedure to the multi-item case. Finally, Section 7 discusses possible extensions, future research directions and conclusions. The electronic companion of this paper containing all proofs and appendices is available together with the online version that can be found at http://or.journal.informs.org/.

2   Problem Formulation Single-Item Model

As explained above, the system under consideration is a centralized continuous review inventory system with one warehouse and N non-identical retailers. Initially (in Sections 2, 3 and 4), we limit

our attention to single-item systems. Extensions to multi-item systems are described in Section 6. The warehouse replenishes its inventory from one or several outside suppliers with constant lead time L_0. The retailers order solely from the central warehouse (no lateral transshipments between retailers are allowed). All stockpoints use complete backordering and FCFS allocation of items to demands. Customer demand occurs at the retailers and follows stationary compound Poisson processes with discrete compounding distributions. Thus, customer orders at retailer i arrive according to a Poission process, and the size of each order is random with a discrete probability mass function (referred to as a compounding distribution). The demand processes across different retailers can be different but they are assumed to be mutually independent. An important rational for using the compound Poisson process is that it offers the flexibility to model highly variable customer demand processes with order size uncertainty. This is a characterizing feature of the industry example motivating this work, but it also applies to many other real world systems (e.g. Kapuscinski et al.(2004)). For practical examples of using compound Poisson processes to model real world demand we refer to, for example, Eaves (2002), Berling and Marklund (2013) and Lengu et al. (2014). Theoretically, the compound Poisson process is interesting because of its generality. For instance, it can be shown (Feller 1966, Axsäter 2006) that any demand process, where the cumulative demand follows a non-decreasing stochastic process with mutually independent increments (a common assumption in the inventory literature), can be represented by a limit of an appropriate sequence of compound Poisson processes.

Inventory control at the retailers is accomplished by base-stock policies with order-up-to level S_i for retailer i. The use of base-stock policies is not an assumption per se, but a consequence of the assumption that POS (Point Of Sale) information is immediately transferred to the warehouse without any fixed costs associated with each transaction. Thus, there are no incentives to batch demand information from consecutive customers into consolidated orders. However, there are fixed costs associated with the transportation of physical products. These costs are considered centrally in the shipment consolidation of deliveries from the warehouse to the retailers over time and across groups of retailers sharing common delivery routes. In accordance with the referred literature on continuous review multi-echelon inventory systems all stockpoints apply a partial delivery policy. At the retailers, this means that a customer that orders more than what is on hand at the retailer will receive the units on hand and the remaining demand is backordered. Similarly at the warehouse, if only part of a retailer order (corresponding to a customer order) is on hand when a shipment is about to leave, the available units will be shipped immediately and the rest later. Partial deliveries

are commonly used in practice (for instance, in the industry examples motivating this work) as it represents the fastest way to satisfy the customer demand. It dominates the obvious alternative of only delivering complete orders provided that there are no, or very small, costs associated with order splitting. This holds for example in a distribution system where there are only variable shipment costs (including picking, loading transporting and receiving) per unit handled. Or as in our case with periodic shipments, where there are variable and fixed shipment costs, but the latter are associated with the planned shipments, which are unaffected by order splitting. If splitting an order leads to an extra shipment associated with fixed costs, a partial delivery policy may be costly. Our method can be modified to deal with complete instead of partial deliveries analogously to the analysis in Howard and Stenius (2013).

The central warehouse uses a continuous review (R_0, nQ_0) policy to control its inventory replenishments. This means that as soon as the (installation stock) inventory position (defined below) falls to or below R_0 an order of nQ_0 units is placed with an outside supplier/manufacturer. Here n is the smallest integer that raises the inventory position above R_0. The batch quantity Q_0 is presumed to be given with the restriction that it must be a positive integer (in our motivating example it is typically determined by set up costs and restrictions at the outside supplier/manufacturer, but it can also be determined by a deterministic EOQ method, as suggested in Zheng 1992 and Axsäter 1996). Even though Q_0 is not a decision variable in our model, the presented method can of course be used repeatedly to evaluate different Q_0 options. Extensions to other replenishment policies at the central warehouse are discussed in Section 7. The use of continuous review base stock policies at the retailers implies that from a control perspective, there is no difference between using an installation stock policy and an echelon-stock policy at the warehouse. Thus the former concept is used in this paper.

The shipment consolidation policy for outbound warehouse deliveries means that there is a shipment leaving the central warehouse to retailer i every T_i time units. Henceforth, T_i is referred to as the shipment interval of retailer i. Periodic shipments are often seen in practice as they facilitate planning of dispatching activities, such as picking, loading, transportation and receiving, and can therefore reduce the costs (see, for example, Gaur and Fisher 2004, and Kuhn and Sternbeck, 2013). In addition to consolidation of outbound deliveries to a single retailer over the shipment intervals, the model allows for consolidation of shipments to groups of retailers sharing the same delivery routes. This is done by synchronizing the shipment intervals (the lengths and moments of shipments) to all retailers on the same route, allowing the same vehicle(s) to service all retailers in that group with a joint shipment.

Note that because of the FCFS allocation, synchronization of shipment intervals across retailer groups is of no relevance in our model. With slight abuse of notation we let T_k denote the shipment interval to retailer group k. The retailer groups and the routing within the groups are taken as given input to our model and need to be determined exogenously. However, the model can of course be used to evaluate different groupings and routing alternatives. In the industry examples motivating this work, restrictions in transport capacity has not been considered an issue as it is bought from third party logistics providers/forwarders. Therefore no such restrictions are included in the model.

The shipment time from the warehouse to retailer i, L_i, includes not only handling activities such as picking, loading, transporting and receiving, but also the time to visit the preceding retailers on the same route. The replenishment lead times for retailer orders are stochastic and depend on the shipment times, the shipment intervals and the stock availability at the warehouse. Note that the latter two affect the delay at the warehouse before a unit is shipped. The replenishment lead time for the central warehouse, L_0, is assumed constant.

To further explain the sequence of events in the replenishment and delivery processes, consider the system as a customer arrives to retailer i at time t, and demands x units. The retailer then tries to satisfy the demand from its stock on-hand, and if there are less than x units on hand, the shortage is backordered. These backorders are cleared in a FCFS sequence as forthcoming replenishments arrive from the warehouse. When the customer arrives at time t, the demand instantaneously translates into an order of x units from retailer i to the warehouse, which then reserves x units for delivery to retailer i. If there is unreserved stock on-hand at the warehouse, these units are reserved first and are added to the reserved stock on-hand awaiting the next shipment to retailer i, W_i. If there is less than x unreserved units on-hand, the remaining units are backordered (reserved units that are outstanding at time t). As replenishments arrive, the backorders are cleared in a FCFS sequence and the units are added to the reserved stock on-hand. When the next shipment is dispatched to retailer i, the reserved stock on-hand for this retailer, W_i, goes to zero as all these units leave the warehouse on a shipment to retailer i. Focusing on the replenishment process at the warehouse, the reservation of x units at time t lowers the warehouse's inventory position, defined as the outstanding orders + unreserved stock on-hand − backorders, by the same amount. When the inventory position falls to or below R_0, an order of nQ_0 units is placed with the outside supplier, bringing the inventory position back between $R_0 + 1$ and $R_0 + Q_0$. Note that the reserved stock on-hand is excluded from the inventory position as these units cannot be used for satisfying future retailer orders. The reserved stock-on hand at the

warehouse can for all practical purposes be seen as units on route to specific retailers.

Analogous to the inventory position above, the inventory level at the central warehouse, IL_0, is defined as the unreserved stock on-hand minus the backorders. Note, if $T_i \to 0 \; \forall \; i = 1, ..., N$ and the warehouse uses partial deliveries to satisfy retailer orders, IL_0 represents the traditional installation-stock inventory level. The total stock on hand at the warehouse is the sum of the unreserved stock on-hand, $max(IL_0, 0)$, and the reserved stock-on hand awaiting shipment to different retailers, $\sum_{i=1,...,N} W_i$. The inventory level of retailer i, IL_i, is defined as the stock on hand minus the amount of backorders at this retailer. The inventory position at retailer i is the inventory level plus all outstanding orders, and it is kept at the base stock level S_i at all times.

The FCFS allocation policy is commonly used in practice. It is generally considered as a "fair" allocation policy that is easy to implement. In the current model, it also provides an incentive for the retailers to share their point-of-sale information immediately with the warehouse. However, it is clearly not an optimal allocation policy. Inventory allocation in divergent multi-echelon systems with stochastic demand is an inherently difficult problem, and no general optimality results exist. FCFS allocation allows tractability and is a standard assumption in the continuous review literature. To our knowledge relaxations are only analyzed exactly in Marklund (2006) and Axsäter and Marklund (2008) both assuming Poisson demand. FCFS allocation is also used in numerous periodic review models of divergent systems starting with Graves (1996) coining the term virtual allocation. Turning to our current model setting, Howard and Marklund (2011) and Howard (2013) have investigated the performance of FCFS allocation in comparison to four state dependent myopic allocation policies (two in each paper) under the assumption of Poisson demand. The two policies in Howard (2013) have a performance guarantee over FCFS. In each paper, the two policies analyzed are based on postponing the allocation to the moment of shipment, and the moment of delivery respectively. The numerical results obtained by simulation show that when allocation is performed at the moment of delivery, the cost savings of using the more sophisticated policies instead of FCFS can be significant: on average 5.6% in Howard and Marklund (2011) and 6.7% in Howard (2013). Allocation at moment of delivery is for practical reasons very challenging to implement and is associated with additional costs for inventory handling and IT, which are not included in the analysis. When allocation is performed at the moment of shipment, which is much cheaper and easier to implement, the savings are quite small: 1.6% and 2.4% respectively. The compound Poisson assumption in our present work implies larger demand variability than in the studied models with Poisson demand. However, there

10

are no indications in these studies that a larger standard deviation to mean ratio of the demand per time unit (i.e. a lower customer arrival rate) has any significant impact on the FCFS performance. Still, further investigation of the performance of FCFS allocation for compound Poisson demand is an interesting venue for future research. In a separate study, Graves (1996) finds a lower bound on the costs in a distribution system with periodic replenishments for any allocation policy. A numerical study indicates that the cost increase of using FCFS (referred to as virtual allocation) is modest. Thus our choice of FCFS allocation is motivated by its wide spread use in practice and in theory, but also by its analytical tractability and seemingly good performance.

We use the following notation to express the system stock levels and demand structure:

N: Number of retailers

K: Number of retailer groups $(\leq N)$

N_k: Set of retailers belonging to retailer group k

λ_i: Arrival rate of customers at retailer i

Y_i: Number of units demanded by an arbitrary customer at retailer i, stochastic variable,

$\quad \mu_i = E[Y_i]$ (Note by assumption $Y_i > 0$ and thus $P(Y_i = 0) = 0$.)

λ_0: Arrival rate of retailer orders at the warehouse

Y_0: Number of units in a retailer order at the warehouse, stochastic variable

λ_{i^c}: Arrival rate of retailer orders at the warehouse excluding orders from retailer i

Y_{i^c}: Number of units in a retailer order at the warehouse excluding orders from retailer i, stochastic variable

$D_i(t_1, t_2)$: Customer demand at retailer i in the time interval $(t_1, t_2]$, where $t_1 \leq t_2$

$D_i(x)$: Total customer demand at retailer i during x units of time (for simplicity $D_i \equiv D_i(1)$)

$D_0(x)$: Aggregate demand at the warehouse during x units of time ($D_0 \equiv D_0(1)$)

$\mathbf{S}$: Vector of retailer order-up-to levels $= (S_1, ..., S_N)$

T_i: Shipment interval to retailer i $(= T_k$ for all retailers $i \in N_k)$

$\mathbf{T}$: Vector of shipment intervals to all retailers $= (T_1, ..., T_N)$

$IL_i(t)$: Inventory level at retailer i at time t $(=$ stock on hand $-$ backorders$)$

$W_i(t)$: Reserved stock on hand at the warehouse destined for retailer i at time t

$IL_0(t)$: Inventory level at the warehouse at time t $(=$ unreserved stock on-hand $-$ backorders$)$

$IP_0(t)$: Inventory position of the warehouse at time t $(IL_0(t) +$ outstanding orders$)$

$x^+ = max(x, 0)$

$x^- = max(-x, 0)$

As the retailer order processes are identical to the customer demand processes, $\lambda_0 = \sum_{i=1}^{N} \lambda_i$ and $Y_0 = \sum_{i=1}^{N} \chi_i Y_i$, where χ_i is an indicator function that is 1 if a given order is from retailer i and 0 otherwise. The probability that a given order emanates from retailer i is $P\{\chi_i = 1, \chi_j = 0 \,\forall\, j \neq i\} = \frac{\lambda_i}{\lambda_0}$ for all $i \in N$. Analogously $\lambda_{i^c} = \sum_{j \in N \setminus i} \lambda_j$ and $Y_{i^c} = \sum_{j \in N \setminus i} \chi_j Y_j$. Based on the definitions above, the probability mass function (pmf) of the demand during x time units at retailer i is

$$P\{D_i(x) = y\} = \sum_{n=1}^{y} \frac{(\lambda_i x)^n}{n!} e^{-\lambda_i x} P\{Y_i^n = y\}, \quad y \geq 0 \tag{1}$$

where Y_i^n represents the total amount of units demanded by n customer orders arriving to retailer i. More precisely, Y_i^n is the sum of n i.i.d. customer orders, Y_i, and the pmf of Y_i^n is determined by the n-fold convolution of the pmf of Y_i. $P\{D_0(x) = y\}$ and Y_0^n are determined analogously.

With respect to cost parameters, the model considers inventory holding costs per unit and time unit at all inventory locations, denoted h_i for $i = 0, 1, ..., N$, and backorder costs per unit and time unit at all retailers, denoted β_i for $i = 1, ..., N$. Moreover, a shipment cost of ω_k is incurred for each scheduled shipment leaving the central warehouse for retailer group k. This reflects, for example, the use of a third party logistics provider or forwarder (as in our motivating industry example) with a fixed cost as part of the contract. Note that ω_k should capture the fixed costs for all retailers in group k, and is incurred even if there are no units to ship. There are also per unit costs for shipping to a specific retailer. However, due to the complete backordering, all demanded units will eventually be shipped. Thus the shipment costs per unit do not affected the optimization and are therefore excluded from the analysis.

The objective is to minimize the long-run average total cost, $TC(R_0, \mathbf{S}, \mathbf{T})$, in (2) with respect to R_0, $\mathbf{S}$ and $\mathbf{T}$. This total cost consists of: (i) The expected warehouse holding cost per time unit, $h_0(E[IL_0^+] + \sum_{i=1,...,N} E[W_i])$, (ii) the expected holding and backorder costs per time unit at retailer i, $h_i E[IL_i^+]$ and $\beta_i E[IL_i^-]$, and (iii) the expected shipment cost per time unit $SC(\mathbf{T}) = \sum_{k=1,...,K} \frac{\omega_k}{T_k}$. Note that the latter is independent of R_0 and $\mathbf{S}$ and only depends on the shipment intervals.

$$TC(R_0, \mathbf{S}, \mathbf{T}) \quad = h_0(E[IL_0^+] + \sum_{i=1,...,N} E[W_i]) + \sum_{i=1,...,N} \left(h_i E[IL_i^+] + \beta_i E[IL_i^-] \right) + \sum_{k=1,...,K} \frac{\omega_k}{T_k}. \tag{2}$$

In a system with fill rate constraints, the objective is to minimize the total costs while meeting the fill rate constraint for each retailer. In this case there will be no backorder costs at the retailers and the total cost function will only consist of holding and shipment costs, that is, the term $\beta_i E[IL_i^-]$ disappears from (2). The fill rate is defined as the portion of the total demand that can be satisfied from stock on hand. How to analyze the fill rates is described at the end of Section 3.2.

3   Analysis

In this section, we provide an exact analysis of the probability mass function of the inventory levels and the expected inventory holding and backorder costs for the single-item system. The analysis is valid for any combination of the decision variables R_0, **S** and **T**. In Section 3.1 we determine the average stock on hand (both unreserved and reserved) at the central warehouse. Section 3.2 explains how to compute the average stock on hand, backorders and fill rates at each retailer i. This analysis assumes that the probability mass functions of the backorders at the central warehouse, designated to retailer $i \; \forall \; i \in N$ are known. Section 3.3 provides an exact approach for determining these probabilities. All proofs are deferred to the online Appendix B, and an illustrative numerical example is available in the online Appendix C. All online appendices are found in the electronic companion.

3.1   The Stock on Hand at the Central Warehouse

As explained above, the stock on hand at the central warehouse consist of unreserved stock on hand, IL_0^+, and reserved stock on hand for each retailer i, W_i. Note that the shipment consolidation policy has no impact on the unreserved stock on hand or the backorders at the central warehouse. The inventory level at the central warehouse, IL_0,(per definition excluding the reserved stock on hand) can therefore be analyzed as a single-echelon system without shipment consolidation. Thus, in steady state $IL_0 = IP_0 - D_0(L_0)$. Moreover, because of the centralized access to POS information (manifested by the continuous review base-stock policies at the retailers) the order process at the warehouse is a superposition of the compound Poisson demand process at the retailers. Assuming that not all customer demand sizes are multiples of some integer larger than one, the inventory position in steady state is uniform on $[R_0 + 1, R_0 + Q_0]$ (see Axsäter 2006 p.88). It follows that

$$E[IL_0^+] = \frac{1}{Q_0} \sum_{IP_0=R_0+1}^{R_0+Q_0} E[(IP_0 - D_0(L_0))^+]. \tag{3}$$

The expected reserved stock on hand destined for retailer i, W_i, is given in Proposition 1.

Proposition 1. *The expected stock on hand at the central warehouse reserved for retailer i is*

$$E[W_i] = \frac{1}{2}\lambda_i\mu_i T_i. \tag{4}$$

3.2 Inventory Levels at the Retailers

To derive the probability mass function for the inventory level at a given retailer, consider a shipment destined for this retailer (group) that leaves the warehouse at time t_0. This shipment arrives at retailer i at time $t_0 + L_i$ and the next shipment will arrive at $t_0 + L_i + T_i$. We call this recurring time interval, $(t_0 + L_i, t_0 + L_i + T_i]$, a replenishment cycle for retailer i (see Figure 1).
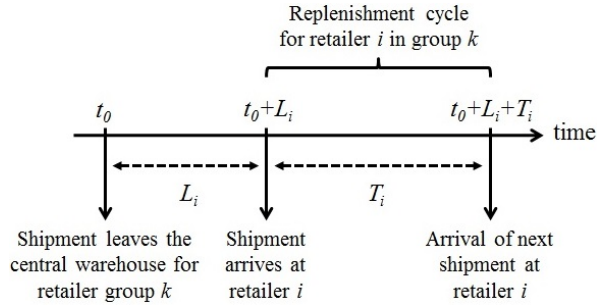
Figure 1: Replenishment cycle for retailer i

The inventory level for retailer i at $t_0 + L_i + t$ $(0 < t \le T_i)$ is determined by the inventory level at $t_0 + L_i$, just after the considered shipment has arrived, and the demand in $(t_0 + L_i, t_0 + L_i + t]$. Defining $B_i(t_0)$ as the number of backordered units for retailer i at the central warehouse at t_0,

$$
\begin{aligned}
IL_i(t_0 + L_i + t) &= IL_i(t_0 + L_i) - D_i(t_0 + L_i, t_0 + L_i + t) \\
&= S_i - B_i(t_0) - D_i(t_0, t_0 + L_i + t), \quad 0 < t \le T_i.
\end{aligned}
\tag{5}
$$

Note that $B_i(t_0)$ and $D_i(t_0, t_0 + L_i + t)$ are independent since $B_i(t_0)$ depends on the demand before time t_0. The pmf of the inventory level of retailer i at $t_0 + L_i + t$ can then be obtained as

$$P\{IL_i(t_0 + L_i + t) = j\} = \sum_{r=0}^{S_i-j} P\{B_i(t_0) = r\}P\{D_i(L_i + t) = S_i - j - r\}, \quad j \le S_i. \tag{6}$$

The challenging part in (6) is to determine the probabilities $P\{B_i(t_0) = r\}$. Section 3.3 explains how this can be done. For now we assume that these probabilities are known and focus on determining

the expected stock on hand, and the expected backorders at retailer i. The former is obtained as the expected stock on hand during a replenishment cycle divided by the cycle length, see (7).

$$
\begin{aligned}
E[IL_i^+] &= \frac{1}{T_i} \int_0^{T_i} E[IL_i(t_0 + L_i + x)^+] dx \\
&= \frac{1}{T_i} \sum_{j=1}^{S_i} j \int_0^{T_i} P\{IL_i(t_0 + L_i + x) = j\} dx \\
&= \frac{1}{T_i} \sum_{j=1}^{S_i} \sum_{r=0}^{S_i-j} j P\{B_i(t_0) = r\} \int_0^{T_i} P\{D_i(L_i + x) = S_i - j - r\} dx
\end{aligned}
\tag{7}
$$

The expected inventory level can be determined analogously,

$$
E[IL_i] = S_i - E[B_i(t_0)] - \lambda_i \mu_i \left(L_i + \frac{T_i}{2} \right).
\tag{8}
$$

Based on these results, the expected number of backorders at retailer i can be obtained from (9).

$$
E[IL_i^-] = E[IL_i^+] - E[IL_i]
\tag{9}
$$

The fill rate for retailer i, denoted by γ_i, can then be determined from (10). This expression follows from analysis of single-echelon systems with compound Poisson demand, see, for example, Axsäter (2006). The difference between the stock on hand at the beginning and at the end of the replenishment cycle renders the amount of demand satisfied directly from stock. This amount is divided by the total expected demand during a cycle. Recall that in systems with fill rate constraints we assume that there are no backorder costs, and the total cost consists of the sum of the holding and shipment costs.

$$
\gamma_i = \frac{E[IL_i(t_0 + L_i)^+] - E[(IL_i(t_0 + L_i) - D(t_0 + L_i, t_0 + L_i + T_i))^+]}{\lambda_i \mu_i T_i}
\tag{10}
$$

3.3 Distribution of the Warehouse Backorders

Consider the central warehouse at time t_0 when a shipment is leaving the central warehouse. We are interested in the pmf of $B_i(t_0)$, the number of backordered units destined for retailer i. The analysis focuses on the inventory level at the warehouse, IL_0 (where all the backorders at the central warehouse are included). The backorders at time t_0 depend on the inventory position a replenishment lead time earlier, $IP_0(t_0 - L_0)$, and the demand during the lead time. Since the inventory position in steady

state is uniformly distributed on $[R_0 + 1, R_0 + Q_0]$, the pmf of $B_i(t_0)$ can be obtained as

$$P\{B_i(t_0) = r\} = \frac{1}{Q_0} \sum_{S_0=R_0+1}^{R_0+Q_0} P\{B_i(t_0) = r | IP_0(t_0 - L_0) = S_0\} \tag{11}$$

The remaining analysis focuses on deriving expressions for $P\{B_i(t_0) = r | IP_0(t_0 - L_0) = S_0\}$ and is divided in two cases: $S_0 > 0$ (in Section 3.3.1) and $S_0 \leq 0$ (in Section 3.3.2).

3.3.1 The Case of $S_0 > 0$

Note first that backorders at time t_0 can only occur when the demand during the replenishment lead time, $D_0(L_0)$, is larger than S_0. Thus, the total amount of units backordered at time t_0, $B_0(t_0)$, is

$$B_0(t_0) = (D_0(L_0) - S_0)^+. \tag{12}$$

Because of the FCFS policy, it will always be the last units demanded in $(t_0 - L_0, t_0]$ that are backordered. To track these units, we therefore study the retailer orders during $(t_0 - L_0, t_0]$. We define:

Φ_0 : Total number of retailer orders arriving to the central warehouse (i.e., number of customers arriving to the system) during $(t_0 - L_0, t_0]$.

Ψ^n: The *nominal inventory position* = the inventory position at the central warehouse at time $t_0 - L_0$ minus the aggregate demand of the n first retailer orders after time $t_0 - L_0$, Y_0^n,

$$\Psi^n = S_0 - Y_0^n. \tag{13}$$

The nominal inventory position helps us to track which retailer orders are backordered at time t_0. Note that $\Psi^0 = S_0$ and $\Psi^{\Phi_0} = IL_0(t_0)$, but for all other values $0 < n < \Phi_0$, Ψ^n can neither be seen as the inventory position nor the inventory level.

For the analysis we divide the backorders in two categories; partial backorders, $\hat{B}_0(t_0)$, and complete backorders, $\check{B}_0(t_0)$, such that $B_0(t_0) = \hat{B}_0(t_0) + \check{B}_0(t_0)$. The partial backorders (at time t_0) result from a retailer order that brings the nominal inventory position from a strictly positive value to a non-positive value. Hence, there can be at most one retailer order that is partially backordered, but it may involve many units that are all referred to as partial backorders. All orders after the nominal inventory position has reached zero will be completely backordered and all these units are referred to

as complete backorders, see Figure 2(a).
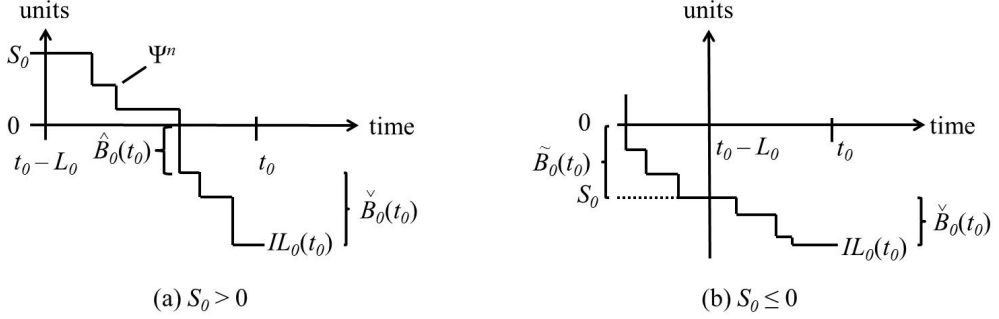
Figure 2: Illustration of two possible sample paths for the nominal inventory position, Ψ^n, at the central warehouse and the associated number of partial and complete backorders

Starting with the analysis of the partial backorders for retailer i we define

$\hat{B}_i^n(t_0)$: Number of partial backorders for retailer i at t_0, when the nth retailer order after $t_0 - L_0$ brings the nominal inventory position to a non-positive value.

Lemma 1. *The probability of u partial backorders for retailer i at time t_0, when the nth customer order after $t_0 - L_0$ brings the nominal inventory position to 0 or below, $P\{\hat{B}_i^n(t_0) = u\}$, is for $u > 0$*

$$P\{\hat{B}_i^n(t_0) = u\} = P\{\Phi_0 \geq n\} \sum_{x=1}^{S_0-n+1} P\{\Psi^{n-1} = x\} \frac{\lambda_i}{\lambda_0} P\{Y_i = u + x\}, \tag{14}$$

and for $u = 0$

$$P\{\hat{B}_i^n(t_0) = 0\} = P\{\Phi_0 \geq n\} \sum_{x=1}^{S_0-n+1} P\{\Psi^{n-1} = x\} \left(\frac{\lambda_i}{\lambda_0} P\{Y_i = x\} + \frac{\lambda_{i^c}}{\lambda_0} P\{Y_{i^c} \geq x\} \right). \tag{15}$$

Turning to the retailer orders that are completely backordered, we know that after the nominal inventory position has reached zero for the nth retailer order, all subsequent orders are classified as complete backorders. Consequently, the distribution of the complete backorders is dependent on n.

$\check{B}_{i,n}(t_0)$: Number of complete backorders for retailer i at t_0, given that the nominal inventory position is taken from a positive to a non-positive value by the nth retailer order after time $t_0 - L_0$ and $\Phi_0 \geq n$.

$\check{\Phi}_{i,n}$: Number of orders from retailer i before t_0 but after the nth retailer order has arrived to the warehouse, given $\Phi_0 \geq n$.

17

Z_i^m : number of orders from retailer i given m orders in total.

Note that Z_i^m is binomially distributed due to the Poisson arrivals of customer orders:

$$P\{Z_i^m = a\} = \binom{m}{a} \left(\frac{\lambda_i}{\lambda_0}\right)^a \left(\frac{\lambda_0 - \lambda_i}{\lambda_0}\right)^{m-a}. \tag{16}$$

Lemma 2. *The probability of v complete backorders for retailer i given that the nth retailer order after $t_0 - L_0$ brings Ψ^n to a non-positive value, $P\{\check{B}_{i,n}(t_0) = v\}$, can be obtained as*

$$P\{\check{B}_{i,n}(t_0) = v\} = \sum_{a=1}^{v} P\{\check{\Phi}_{i,n} = a\}P\{Y_i^a = v\}, \tag{17}$$

where

$$P\{\check{\Phi}_{i,n} = a\} = \sum_{m=a}^{\infty} \frac{P\{\Phi_0 = n + m\}}{P\{\Phi_0 \geq n\}} P\{Z_i^m = a\} \tag{18}$$

Denoting the probability that the nominal inventory position never reaches zero by p_0, we have

$$p_0 = \sum_{n=0}^{S_0-1} P\{\Phi_0 = n\}P\{\Psi^n > 0\}. \tag{19}$$

The pmf of the number of backordered units for $S_0 > 0$ can now be obtained from Proposition 2.

Proposition 2. *The probability that the central warehouse has r backordered units allocated to retailer i at t_0 when the inventory position is $S_0 > 0$ can be obtained for $r > 0$ as*

$$P\{B_i(t_0) = r|IP_0(t_0 - L_0) = S_0\} = \sum_{n=1}^{S_0} \sum_{u=0}^{r} P\{\hat{B}_i^n(t_0) = u\}P\{\check{B}_{i,n}(t_0) = r - u\}, \tag{20}$$

and for $r = 0$ as

$$P\{B_i(t_0) = 0|IP_0(t_0 - L_0) = S_0\} = p_0 + \sum_{n=1}^{S_0} P\{\hat{B}_i^n(t_0) = 0\}P\{\check{B}_{i,n}(t_0) = 0\}. \tag{21}$$

Remark. The expected amount of backorders is not always proportional to the demand per time unit resulting in different average waiting times per unit due to stockouts across the retailers. The expected amount of complete backorders are proportional to the demand per time unit, but this is not always true for the expected amount of partial backorders.

3.3.2 The Case of $S_0 \leq 0$

When the initial inventory position is less than or equal to zero, all units ordered during the time interval $(t_0 - L_0, t_0]$ will be completely backordered at time t_0. We denote this part of the backorders $\check{B}_0(t_0)$ and the units ordered by retailer i, $\check{B}_i(t_0)$. In addition to this, the last $-S_0$ units ordered before time $t_0 - L_0$ will also be backordered at time t_0. We denote this part of the backorders as $\tilde{B}_0(t_0)$ and the units ordered by retailer i $\tilde{B}_i(t_0)$, see Figure 2(b). It follows that $B_0(t_0) = \check{B}_0(t_0) + \tilde{B}_0(t_0)$, $B_i(t_0) = \check{B}_i(t_0) + \tilde{B}_i(t_0)$ and $B_0(t_0) = \sum_{i=1}^{N} B_i(t_0)$. Note that $\check{B}_0(t_0)$ only consists of units from completely backordered retailer orders (ordered after $t_0 - L_0$), while $\tilde{B}_0(t_0)$ $(= -S_0)$ can consist of units both from completely backordered retailer orders, and from a partially backordered retailer order. As $\tilde{B}_i(t_0)$ depends on the demand before time $t_0 - L_0$, and $\check{B}_i(t_0)$ depends on the demand during $(t_0 - L_0, t_0]$, they are independent. The probability of r units backordered at the warehouse for retailer i at time t_0, when $S_0 \leq 0$ is thus

$$P\{B_i(t_0) = r | IP_0(t_0 - L_0) = S_0\} = \sum_{u=0}^{min(r, -S_0)} P\{\tilde{B}_i(t_0) = u\} P\{\check{B}_i(t_0) = r - u\}, \quad S_0 \leq 0. \quad (22)$$

Because all units ordered in time interval $(t_0 - L_0, t_0]$ will be completely backordered, the probability that v units are backordered for retailer i at t_0 is simply

$$P\{\check{B}_i(t_0) = v\} = P\{D_i(t_0 - L_0, t_0) = v\}. \quad (23)$$

Turning to the analysis of $\tilde{B}_i(t_0)$, note first that for $S_0 = 0$, there can be no backorders ordered before $t_0 - L_0$. Consequently $\tilde{B}_i(t_0) = 0 \; \forall \; i$, which means that $P\{\tilde{B}_i(t_0) = 0\} = 1 \; \forall \; i$ and (22) simplifies to $P\{B_i(t_0) = r | IP_0(t_0 - L_0) = 0\} = P\{\check{B}_i(t_0) = r\}$. In order to determine $\tilde{B}_i(t_0)$ for $S_0 < 0$ we study the system backwards in time from $t_0 - L_0$. We define

Ψ^{-m} : Nominal inventory position before $t_0 - L_0$, defined as the inventory position at time $t_0 - L_0$ plus the accumulated demand of the last m retailer orders before $t_0 - L_0$.

$\tilde{B}_i^{-m}(t_0)$: Backordered units to retailer i at t_0 ordered before $t_0 - L_0$, when the nominal inventory position reaches a non-negative value by the occurrence of the mth retailer order before $t_0 - L_0$ (i.e., counting backwards from $t_0 - L_0$, the mth retailer order is the first order that is backordered; either completely or partially).

V_i^m : The number of units ordered by retailer i, given that a total of m retailer orders have occurred.

Because the nominal inventory position can reach zero only once, $\tilde{B}_i^{-m}(t_0)$ for $m \in [1, -S_0]$ represent mutually exclusive events. Thus, we get

$$
P\{\tilde{B}_i(t_0) = u\} = \begin{cases} \sum_{m=1}^{S_0} P\{\tilde{B}_i^{-m}(t_0) = u\}, & S_0 < 0 \\ 1, & u = 0 \text{ and } S_0 = 0 \\ 0, & \text{otherwise} \end{cases} \tag{24}
$$

where $P\{\tilde{B}_i^{-m}(t_0) = u\}$ can be determined from Lemma 3.

Lemma 3. *The probability for u backordered units at t_0, ordered by retailer i before time $t_0 - L_0$, when the mth customer order before $t_0 - L_0$ is the first to be backordered, is*

$$
P\{\tilde{B}_i^{-m} = u\} = \sum_{x=S_0+m-1}^{-1} \frac{\lambda_{ic}}{\lambda_0} P\{Y_{i^c} \geq -x\} P\{V_i^{m-1} = u \text{ and } \Psi^{-(m-1)} = x\}
$$
$$
+ \frac{\lambda_i}{\lambda_0} P\{Y_i \geq -x\} P\{V_i^{m-1} = u + x \text{ and } \Psi^{-(m-1)} = x\}, \tag{25}
$$

where

$$
P\{V_i^{m-1} = u \text{ and } \Psi^{-(m-1)} = x\} = \sum_{a=0}^{m-1} P\{Z_i^{m-1} = a\} P\{Y_i^a = u\} P\{Y_{i^c}^{m-1-a} = x - S_0 - u\}. \tag{26}
$$

$P\{B_i(t_0) = r | IP_0(t_0 - L_0) = S_0\}$ for $S_0 \leq 0$ follows from (22), (23), (24) and Lemma 3.

4 Optimization

In this Section we present a method for optimizing the system parameters R_0, **S** and **T** both in systems with backorder costs, and in systems with fill rate constraints. We assume, as before, that the order quantity Q_0 is given by the outside supplier/manufacturer. The method is explained for the single-item case and is extended to the multi-item case in Section 6. The objective is to minimize the total cost function (2), or the sum of expected shipment and holding costs subject to fill rate constraints. We know from Section 3.1 that the stock on hand, and therefore also the holding costs, at the warehouse can be separated into unreserved stock on hand, IL_0^+, which depends on R_0 but is independent of **T** and **S**, and the reserved stock on hand that depends on **T** but is independent of R_0 and **S**. We define the costs directly related to retailer group k (including the holding cost of the

reserved stock on hand going to the retailers in the retailer group) as

$$TC_k(R_0, \mathbf{S}, T_k) = \frac{\omega_k}{T_k} + \sum_{i \in N_k} h_0 E[W_i] + h_i E[IL_i^+] + \beta_i E[IL_i^-]. \tag{27}$$

Recall that, with slight abuse of notation, we let T_k denote the shipment interval for all retailers in retailer group k ($T_k = T_i \ \forall \ i \in N_k$). The total cost function can thus be expressed as

$$TC(R_0, \mathbf{S}, \mathbf{T}) = h_0 E[IL_0^+] + \sum_{k \in K} TC_k(R_0, \mathbf{S}, T_k). \tag{28}$$

The total cost function is not jointly convex in R_0, $\mathbf{S}$ and $\mathbf{T}$ which can be shown by example. Thus, the proposed optimization method is based on bounding R_0 and $T_k \ \forall \ k$, using Proposition 4 and Proposition 5 respectively, and searching this bounded region using the convexity property of the retailer order-up-to levels specified in Proposition 3. The approach for obtaining the bounds in Propositions 4 and 5 is to assert when further changes of the considered decision variable can no longer decrease the total cost below the lowest cost known so far, $\overline{TC}$.

Proposition 3. *For fixed R_0 and $\mathbf{T}$ the total cost function $TC(R_0, \mathbf{S}, \mathbf{T})$ is convex and separable in the retailer order-up-to levels $\mathbf{S}$.*

4.1 Optimization procedure

The first step in the optimization procedure is to determine lower bounds for the optimal total costs directly related to each retailer group $k \in K$, TC_k^l valid for all $\mathbf{T}$, R_0 and $\mathbf{S}$. Explanations of how TC_k^l $\forall \ k$ can be obtained are given in Section 4.2. The second step is to define $\overline{TC}$ as the lowest total cost under all currently known policies, and use a heuristic to determine a good initial value for $\overline{TC}$ (a close to optimal initial solution will provide tighter bounds). The heuristic solution is determined by first fixing the shipment intervals to some near optimal values $\overline{\mathbf{T}}$, i.e. $T_k = \overline{T}_k \ \forall \ k$, and then optimizing R_0 and $\mathbf{S}$. Values for $\overline{T}_k \ \forall \ k$ can be obtained from the deterministic Economic Order Interval (EOI) heuristic, presented in Marklund (2011). It does not allow for shortages and assumes that the demand per time unit at retailer i is constant and equal to $\lambda_i \mu_i$. Noting that the shipment interval affects the inventory both at the warehouse and at the retailers, the expected cost per time unit for retailer group

k can be expressed as $\sum_{i \in N_k}(h_0 + h_i)\overline{T}_k \frac{\lambda_i \mu_i}{2} + \frac{\omega_k}{\overline{T}_k}$. Minimizing this cost with respect to $\overline{T}_k$ renders

$$\overline{T}_k = \sqrt{\frac{2\omega_k}{\sum_{i \in N_k}(h_0 + h_i)\lambda_i \mu_i}}. \tag{29}$$

Given the shipment intervals $\overline{T}_k \ \forall \ k$, R_0 and $\mathbf{S}$ are optimized by searching the possible values of R_0 starting from $R_0^l = -Q_0$ (this bound is known from previous research, see for example Marklund 2011 and Axsäter 1998), and using the convexity property in Proposition 3 to optimize $S_i \ \forall \ i$ for each value of R_0. The search is limited by the upper bound R_0^u in Proposition 4. Note that $\overline{TC}$ and thereby R_0 are updated during the search as better solutions with lower expected costs are found.

Proposition 4. *An upper bound for the optimal reorder point at the central warehouse, R_0^u, is obtained for the smallest value of R_0 satisfying the condition*

$$h_0 E[IL_0^+] \geq \overline{TC} - \sum_{k \in K} TC_k^l. \tag{30}$$

For systems with fill rate constraints at the retailers, the optimization of R_0 and $\mathbf{S}$ (given $\mathbf{T}$) is achieved by searching the interval $[R_0^l, R_0^u]$, and for each R_0 determine the smallest $S_i \ \forall \ i$ satisfying the fill rate constraint at retailer i, utilizing that the expected holding costs are increasing in S_i.

The third step in the optimization procedure is to determine lower and upper bounds for the optimal $T_k \ \forall \ k \in K$, using Proposition 5.

Proposition 5. *For the optimal shipment interval of retailer group k, a lower bound is obtained by*

$$T_k^l = \frac{\overline{TC} - \sum_{\kappa \neq k} TC_\kappa^l - \sqrt{(\overline{TC} - \sum_{\kappa \neq k} TC_\kappa^l)^2 - 2\omega_k \sum_{i \in N_k} h_0 \lambda_i \mu_i}}{\sum_{i \in N_k} h_0 \lambda_i \mu_i}, \tag{31}$$

and an upper bound by

$$T_k^u = \frac{\overline{TC} - \sum_{\kappa \neq k} TC_\kappa^l + \sqrt{(\overline{TC} - \sum_{\kappa \neq k} TC_\kappa^l)^2 - 2\omega_k \sum_{i \in N_k} h_0 \lambda_i \mu_i}}{\sum_{i \in N_k} h_0 \lambda_i \mu_i}. \tag{32}$$

Finally, the optimal solution with respect to $\mathbf{T}$, R_0 and $\mathbf{S}$ is obtained by searching all combinations of T_k within the bounded region, (using some step size τ) and optimizing R_0 and $\mathbf{S}$ for every combination according to the same procedure as for the initial solution (where $T_k = \overline{T}_k \ \forall \ k$) explained above. As better solutions with lower total expected costs are found, $\overline{TC}$ is updated and the bounds on the shipment intervals and R_0 are tightened. When the search is complete, the optimal solution is found

(the optimal system parameters for T_k $\forall k$, R_0 and S_i $\forall i$ are denoted T_k^*, R_0^* and S_i^* respectively) and the associated minimum expected cost, TC^*, equals $\overline{TC}$. Note that the probability mass functions of the warehouse backorders, which can be time consuming to calculate, are independent of T_k and $\mathbf{S}$ and only need to be determined once for each $R_0 \in [R_0^l, R_0^u]$. Moreover, as $\overline{TC}$ is updated during the optimization R_0^u never increases. Thus, the probability mass functions of the warehouse backorders for the initial solution is sufficient for the entire optimization.

4.2 Lower Bound for Costs Directly Related to Retailer Group k

The presented bounds on T_k and R_0 are based on the existence of a lower bound for all costs directly related to retailer group k, TC_k^l, valid for all values of R_0 and T_k. Lemma 4 provides such a bound which is applicable both for systems with backorder costs and fill rate constraints.

Lemma 4.

$$TC_k^l = \sqrt{2h_0\omega_k \sum_{i \in N_k} \lambda_i \mu_i}. \tag{33}$$

The optimization of the retailer order-up-to levels is more time consuming for systems with backorder costs than for systems with fill rate constraints, as a numerical integration is needed when evaluating the expected retailer costs for each value of S_i. To compensate for this, Lemma 5 provides tighter bounds for TC_k^l in backorder cost systems.

Lemma 5. *A lower bound for the costs directly related to retailer group k, $TC_k(R_0, \mathbf{S}, T_k)$ for all R_0, can for systems with backorder costs be obtained by minimizing these costs with respect to T_k and S_i for a system where there are no backorders at the central warehouse ($R_0 \to \infty$):*

$$TC_k^l = \min_{T_k, S_i \forall\ i \in N_k} [TC_k | B_i(t_0) = 0, \forall\ i \in N_k] \le TC_k(R_0, \mathbf{S}, T_k), \forall\ R_0, \mathbf{S}, T_k. \tag{34}$$

$TC_k^l = \min_{T_k, S_i}(TC_k | B_i(t_0) = 0, \forall\ i)$, may be computed by the algorithm in the online Appendix A.

5 Numerical Results

This section summarizes some of the results from a numerical study detailed in Appendix D found in the electronic companion to this paper. The purposes of this study are to investigate: the behavior of the optimal solutions, investigate the quality of the optimality bounds, and the performance of the

EOI heuristic for determining the shipment intervals. This section focuses on results concerning the latter two. It also includes a discussion about computational issues. The study covers 128 problems representing all combinations of a high and a low value for the parameters N, L_i, ρ_i (variance to mean ratio of the demand per time unit), β_i, ω_k, L_0 and Q_0. More precisely, the number of retailers, $N = 3$ or 6; where in both cases we have two retailer groups. When $N = 3$, retailers $\{1, 2\}$ constitute one group and retailer 3 a second. When $N = 6$, retailers $\{1,2,3,4\}$ constitute one group and retailers $\{5, 6\}$ a second. The expected demand per time unit, $E[D_i]$, is $\{2,1,3\}$ when $N = 3$, and $\{2,1,2,1,4,2\}$ when $N = 6$. The shipment times, $L_i = \{1, 2, 1\}$ or $\{2, 4, 2\}$ when $N = 3$, and $\{1, 1, 2, 2, 1, 2\}$ or $\{2, 2, 4, 4, 2, 4\}$ when $N = 6$. We also consider $\rho_i = 1$ (Poisson demand) or 5 (compound Poisson demand with logarithmic compounding distributions) $\forall\ i \in N$, $\beta_i = 10$ or $100\ \forall\ i \in N$, $\omega_k = 10$ or $100\ \forall\ k \in K$, $L_0 = 1$ or 5, $Q_0 = 2$ or 20, and $h_i = 1\ \forall\ i$.

For every problem, $T_k\ \forall\ k \in K$, R_0 and $S_i\ \forall\ i \in N$ are optimized using the method in Section 4 (with step size $\tau = 0.01$ time units for $T_k\ \forall k$). For the optimal solution we compute the distance to the optimality bounds, $\Delta R_0^u = R_0^u - R_0^*$, $\Delta T_k^l = T_k^* - T_k^l$ and $\Delta T_k^u = T_k^u - T_k^*$. We also determine the relative difference between optimal shipment intervals and those obtained by the EOI heuristic, $\Delta T_k = (\overline{T}_k - T_k^*)/T_k^*$, and the relative increase in the associated costs, $\Delta C = (\overline{TC} - TC^*)/TC^*$. The average, maximum and minimum results across all problems are presented in Table 1. From there we can see that the upper bound, R_0^u, which is the most important bound computationally, is rather tight; on average 7.41 units above R_0^*. The importance of R_0^u stems from the fact that the warehouse backorder distributions are the most time consuming to compute (especially when $R_0 + Q_0$ is large). Fortunately, they only need to be computed once for each R_0.

Table 1: Summary results for the numerical study

	$E[TC^*]$	$E[R_0^*]$	$E[S_i^*]$	$E[T_k^*]$	$E[\Delta C]$	$E[\Delta T]$	$E[\Delta R_0^u]$	$E[\Delta T_k^l]$	$E[\Delta T_k^u]$
Average	111.51	19.89	19.41	3.15	0.14%	2.99%	7.41	2.19	23.55
Minimum	31.61	-10	7.67	1.09	0.00%	-9.79%	0	0.91	7.05
Maximum	264.19	68	38.33	6.11	0.66%	18.60%	24	4.23	51.26

Looking at the bounds for T_k, Table 1 shows that T_k^l is on average 2.19 time units below the optimal value T_k^*, while T_k^u exceeds it with on average 23.55 time units. Thus, the lower bound tends to be rather close to the optimum, while the upper bound is looser. Fortunately, from a computational perspective it is less important to provide tight bounds on T_k^* than on R_0^*.

Table 1 indicates that the EOI heuristic performs well and offers a viable option to the optimal solution. On average it overestimates the shipment intervals by 2.99%, but the average cost increase

is only 0.14% with a maximum of 0.66%. This suggests that the total cost is insensitive to the choice of shipment intervals around the optimum as long as R_0 and $\mathbf{S}$ are adjusted accordingly.

The computational complexity of the presented method is directly related to the number of combinations that need to be considered in the cost and fill rate evaluations (follows from the analysis in Section 3 and 4). Consequently, evaluation and optimization of large systems, with many retailers, high demand items, and long lead-times can be time consuming. Looking at the computation times for the problems in the numerical study they range from seconds to several hours, with changes in L_0 and N having the biggest impact. On average, the computation times are 19 times longer for the problems with $L_0 = 5$ than for $L_0 = 1$, and 6.5 times longer for problems with $N = 6$ (with $E[D_0] = 12$) than for $N = 3$ (with $E[D_0] = 6$). In contrast, 10 fold increases of β_i, ω_k and Q_0 increase the computation times on average by 26%, 7%, and 30% respectively. To investigate the computational impact of larger expected warehouse demand for fixed N, we solved 64 additional problems with $N = 3$ for which the expected demand per time unit at each retailer was doubled rendering $E[D_0] = 12$ (keeping the compounding distributions and everything else fixed). On average the computation times then increase by a factor 2.2. Thus, the computational times increase significantly with both N and $E[D_0]$. Increasing the number of retailer groups K (for fixed N) increases the number of candidate solutions to search in the optimization, but it only affects the inventory cost and fill rate calculations if the bounds T_k^l and T_k^u change so new T_i - values need to be considered. For given optimality bounds and step sizes τ_k, there are $(R_0^u - R_0^l) \prod_{k=1}^{K} \left((T_k^u - T_k^l)/\tau_k \right)$ combinations of R_0 and $\mathbf{T}$ to consider. This means that the search space (and thereby the computation times) increases rapidly with K if τ_k is small compared to $(T_k^u - T_k^l)$ for all k, but more modestly otherwise.

6 Multi-Item Systems

This section extends the analysis to multi-item systems, where J items are controlled simultaneously. The items are distributed via a central warehouse to N retailers (The model is also applicable to situations where different items are located at different central warehouses and shipments from these warehouses are consolidated). The items may be distributed in any way across the retailers. However, without loss of generality, we assume that each item is sold by at least one of the retailers and each retailer sells at least one of the items. Shipments may be consolidated to K consolidation groups across items and retailers. A consolidation group consists of a number of items jointly distributed to a collection of retailers. In principle there can be different consolidation groups associated with

different items, but $K \leq N \times J$. Note that in the single-item case a consolidation group is equivalent to a retailer group. From a practical perspective, the possibility to consolidate shipments across items as well as retailers increases both the practical applicability and the possible gain. We define:

$R_{0,j}$: reorder point at the central warehouse for item j $S_{i,j}$: order-up-to level at retailer i for item j

$\mathbf{S}_j$: vector with order-up-to levels for all retailers for item j $(S_{1,j}, ..., S_{N,j})$

$T_{i,j}$: shipment interval for item j to retailer i $(= T_k$ for all items and retailers in group $k)$

$\mathbf{T}$: matrix with shipment intervals for all items to all retailers $(T_{1,1}, ..., T_{1,J}; ...; T_{N,1}, ..., T_{N,J})$

$HC_{0,j}(R_{0,j}, \mathbf{T})$: holding costs at central warehouse for item j

$HC_{i,j}(R_{0,j}, \mathbf{S}_j, \mathbf{T})$: holding costs at retailer i for item j

$BC_{i,j}(R_{0,j}, \mathbf{S}_j, \mathbf{T})$: backorder costs at retailer i for item j

TC_M : the total cost function in the multi-item system

As explained in Section 2, the fixed shipment costs are incurred for every scheduled shipment that leave the warehouse. In the multi-item case, the fixed cost for every scheduled shipment to consolidation group k, ω_k, may, for example, be determined as the sum of the fixed shipment costs per retailer and item belonging to this group. This way to model the costs will assure that consolidation can be obtained both across retailers and items in group k by using the same shipment intervals, and by synchronizing their start. In the cost functions for multi-item systems, the total shipment costs per time unit can still be determined as $SC(\mathbf{T}) = \sum_{k \in K} \frac{\omega_k}{T_k}$.

Given $\mathbf{T}$, the holding and backorder costs as well as the fill rates for item j are independent of the other items and can be determined as in the single-item case. The total cost function for the multi-item system with backorder costs is

$$TC_M = \sum_{j \in J} \left[HC_{0,j}(R_{0,j}\mathbf{T}) + \sum_{i \in N} [HC_{i,j}(R_{0,j}, \mathbf{S}_j, \mathbf{T}) + BC_{i,j}(R_{0,j}, \mathbf{S}_j, \mathbf{T})] \right] + SC(\mathbf{T}). \quad (35)$$

For a multi-item system with fill rate constraints, each item j at retailer i may have an individual fill rate constraint while the backorder costs are excluded from the total cost function. The fill rate of item j at retailer i, $\gamma_{i,j}$, can be determined analogously to the single-item case using (10).

The optimization of multi-item systems is similar to the single-item optimization but with the coupling constraint that all items in a consolidation group use the same shipment interval. The shipment intervals are the only way in which the different items affect each other. The bounds for the

26

warehouse reorder points need to be determined separately for each item, while T_k^l and T_k^u are only determined once for each consolidation group, taking all items into consideration. Propositions 4 and 5 can still be used to determine these bounds, provided that $\overline{TC}$ is redefined as the lowest total cost for all items under all currently known policies, and TC_k^l is redefined as a lower bound for the total costs of all items related to consolidation group k. TC_k^l can be obtained from Lemma 4 by including the holding costs for the reserved stock on hand of all items in group k.

The optimization procedure is analogous to the single-item case. First, TC_k^l is determined. Second, an initial total cost, $\overline{TC}$, is obtained by optimizing a system where the shipment intervals, $\overline{T}_k \; \forall \; k \in K$, are determined by a straightforward extension of the heuristic in (29). In this extension, the total holding cost is a summation of the holding costs of all items associated with the considered consolidation group. The optimization, given $\overline{T}_k$, is performed separately for each item $j \in J$ according to the single-item procedure. Given $\overline{TC}$, initial values for T_k^l and $T_k^u \; \forall \; k \in K$ can be determined and the bounds are tightened as $\overline{TC}$ is updated. Finally, an optimal solution is found by searching all combinations of shipment intervals within the bounds, optimizing the reorder points and order-up-to levels at all retailers for all items separately, and choosing the solution with the lowest total cost.

7   Summary, Extensions and Future Research

In this paper we have presented a method for exact analysis of the inventory level distributions, fill rates and expected costs in one warehouse multi-retailer inventory systems with time based shipment consolidation and compound Poisson demand. An optimization procedure is also provided based on bounding the optimal decision variables; the warehouse reorder point, the retailer base-stock levels, and the shipment intervals. The method is applicable for both single- and multi-item systems.

The analysis can be extended to other related systems, for example with different replenishment policies at the central warehouse. The same technique for analyzing the inventory levels, costs and fill rates at the retailers (via the backorder distribution at the warehouse) can be used for any warehouse replenishment policy for which the inventory position and lead time distributions are known. One such extension is the use of an (s, S) policy at the warehouse. For this system, the lead time (L_0) is constant and the distribution of the warehouse inventory position can for instance be obtained by the technique in Axsäter (2006, p. 107–109). With this distribution known, the cost analysis of the entire system can be obtained analogously to the approach presented in this paper. Assuming that $S - s$ is fixed, the optimization can be handled analogously as well. Relaxing this assumption offers

interesting challenges for further research.

Another possible extension is to let the warehouse use, what we refer to as, *synchronized* (R, nQ) replenishments. This implies that replenishment orders from the central warehouse are placed only when there is a shipment leaving (to any retailer group) exactly L_0 time units later. To elaborate, consider a situation when the inventory position at the warehouse falls below R_0 at time t_0 and the first shipment from the warehouse after time $t_0 + L_0$ leaves at $t_0 + L_0 + \Delta$. If the replenishment order of Q_0 units is placed at time t_0 (as our policy prescribes) it will arrive Δ time units before it is needed. Under the *synchronized* policy the warehouse delays the order placement Δ time units, and will thereby reduce its inventory with Q_0 units for Δ time units. Because of the periodic shipment schedule, this delay does not affect the service to the retailers. The analysis of the retailers in this system therefore remains the same as in our current model. However, the analyses of the warehouse inventory (both the reserved and unreserved stock on hand) and the optimization bounds are severely complicated and beyond the scope of this paper to solve. Note that, theoretically, the *synchronized* policy offers a performance guarantee over the presented (R, nQ) policy. Intuitively, the difference in performance will depend on how often shipments are dispatched. This is influenced, for example, by the number of retailer/consolidation groups, the length of the shipment intervals and the coordination of dispatching times. Frequent shipments suggests smaller opportunities for savings. From a practical perspective, there are two drawbacks with the *synchronized* policy; (1) it is sensitive to variations in the replenishment lead times, L_0, and (2) the receiving and loading activities of all items at the central warehouse are concentrated to the same time instances. Other possible extensions would be periodic (R, nQ) replenishments, and *synchronized* or periodic (s, S) policies.

The exact approach presented in this paper can be computationally challenging to apply to large systems. Thus, another direction for future research is to consider computationally more efficient methods for cost analysis and optimization. In addition, we believe that our analysis provides a good foundation for future research on both exact methods and accurate heuristics for analyzing similar divergent systems. Other possible future areas of research using extensions of this approach includes N-echelon distribution systems, periodic replenishments to the central warehouse or demand with other distributions than compound Poisson.

Electronic Companion

An electronic companion, containing all proofs and appendices, is available together with the online version of this paper at http://or.journal.informs.org/.

References

Andersson, J., S. Axsäter, J. Marklund. 1998. Decentralized multiechelon inventory control. *Production and Operations Management* **7** 370–386.

Andersson, J., J. Marklund. 2000. Decentralized inventory control in a two-level distribution system. *European Journal of Operational Research* **127** 483–506.

Axsäter, S. 1990. Simple Solution Procedures for a Class of Two-Echelon Inventory Problems. *Operations Research* **38** 64–69.

Axsäter, S. 1993a. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* **41** 777–785.

Axsäter, S. 1993b. Optimization of order-up-to-S policies in two-echelon inventory systems with periodic review. *Naval Research Logistics* **40** 245–253.

Axsäter, S. 1996. Using the deterministic EOQ formula in stochastic inventory control. *Management Science* **42** 830–834.

Axsäter, S. 1997. Simple evaluation of echelon stock (R, Q)-policies for two-level inventory systems. *IIE Trans.* **29** 661–669.

Axsäter, S. 1998. Evaluation of installation stock based (R, Q)-policies for two-level inventory systems with Poisson demand. *Operations Research* **46** 135–145.

Axsäter, S. 2000. Exact analysis of continuous review (R, Q)-policies in two-echelon inventory systems with compound Poisson demand. *Operations Research* **48** 686–696.

Axsäter, S. 2001. A note on stock replenishment and shipment scheduling for vendor-managed inventory systems. *Management Science* **44** 1306–1310.

Axsäter, S. 2003. Supply chain operations: Serial and distribution inventory systems. Handbooks in operations research and management science. Vol.**11**: Supply chain management: Design, coordination and operation, S.C. Graves and A.G. de Kok (Editors), Elsevier, Amsterdam, The Netherlands, 525–559.

Axsäter S. 2006. Inventory Control. Springer: Berlin.

Axsäter S., J. Marklund, E.A. Silver. 2002. Heuristic Methods for Centralized Control of One-warehouse, N-Retailer Inventory Systems. *Manufacturing and Service Operations Management* **4** 75–97.

Axsäter S., J. Marklund. 2008. Optimal "Position Based" Warehouse Ordering in Divergent Two-Echelon Inventory Systems. *Operations Research* **56** 976–991.

Berling, P., J. Marklund. 2013. A model for heuristic coordination of real life distribution inventory systems with lumpy demand. *European Journal of Operational Research* **230** 515–526.

Berling, P., J. Marklund. 2014. Multi-echelon inventory control: A normal demand model for implementation in practice. *International Journal of Production Research* **52** 3331–3347.

Cachon G.P. 1999. Exact Evaluation of Batch-ordering Inventory Policies in Two-echelon Supply Chains with Periodic Review. *Operations Research* **49** 79–98.

Cachon, G., M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science* **46** 1032–1048.

Çetinkaya S., C.Y. Lee. 2000. Optimal outbound dispatch policies: Modeling inventory and cargo capacity. *Management Science* **46** 217–232.

Çetinkaya, S., J.H. Bookbinder. 2003. Stochastic Models for the Dispatch of Consolidated Shipments. *Transportation Research* **37** 747–768.

Çetinkaya, S., E. Tekin, C.Y. Lee. 2008. A Stochastic Model for Joint Inventory and Outbound Shipment Decisions. *IIE Transactions* **40** 324–340.

Chen, F., Y.S. Zheng. 1997. One Warehouse Multi-Retailer Systems with Centralized Stock Information. *Operations Research* **45** 275–287.

Cheung K.L., H.L. Lee. 2002. The Inventory Benefit of Shipment Coordination and Stock Rebalancing in a Supply Chain. *Management Science* **48** 300–306.

Chu L.Y., Z.M. Shen. 2010. A Power-of-Two Ordering Policy for One-Warehouse Multi-retailer Systems with Stochastic Demand. *Operations Research* **58** 492–502.

Diks, E.B., A.G. de Kok. 1998. Optimal control of a divergent multi-echelon inventory system. *European Journal of Operational Research* **111** 75–97.

Diks, E.B., A.G. de Kok. 1999. Computational results for the control of a divergent *N*-echelon inventory system. *International Journal of Production Economics* **59** 327–336.

Eaves A.H.C. 2002. Forecasting for the ordering and stock- holding of consumable spare parts. PhD Thesis, University of Lancaster.

Federgruen, A., P. Zipkin. 1984. Approximations of Dynamic Multilocation Production and Inventory Problems. *Management Science* **30** 69–84.

Federgruen, A. 1993. Centralized Planning Models for Multi-Echelon Inventory Systems under Uncertainty, in *S.C. Graves et al. Eds., Handbooks in OR & MS* **4**, North Holland Amsterdam, pp. 133–173.

Feller, W. 1968. An Introduction to Probability Theory and its Applications, Vol. 1, 3rd edition, John Wiley & Sons, New York.

Forsberg, R. 1995. Optimization of order-up-to-*S* policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research* **81** 143–153.

Forsberg, R. 1997. Exact evaluation of (R, Q)-policies for two-level inventory systems with Poisson demand. *European Journal of Operational Research* **96** 130–138.

Gaur, V., M. L. Fisher, 2004. A periodic inventory routing problem at a supermarket chain. *Operations Research* **32** 1019–1037.

Graves, S.C. 1985. A Multi-Echelon Inventory Model for a Repairable Item with One-for-One Replenishment. *Management Science* **31** 1247–1256.

Graves, S.C. 1996. A multi-echelon inventory model with fixed replenishment intervals. *Management Science* **42** 1–18.

Gürbüz, M., K. Moinzadeh, Y.P. Zhou. 2007. Coordinated replenishment strategies in inventory/distribution systems. *Management Science* **53** 293–307.

Heijden M.C. van der, E.B. Diks, A.G. de Kok, 1997. Stock Allocation in General Multi-Echelon Distribution Systems with (R, S) Order-up-to-policies, *International Journal of Production Economics* **49** 157–174.

Houtum G.J. van, K. Inderfurth, W.H.M. Zijm. 1996. Materials Coordination in Stochastic Multi-Echelon Systems. *European Journal of Operational Research* **95** 1–23.

Howard, C. 2013. New Allocation Policies for Divergent Inventory Systems with Real-time Information and Shipment Consolidation, in *Howard, C., Real-time Allocation Decisions in Multi-echelon Inventory Control*, Doctoral Thesis, Div. of Production Management, Dept. of Industrial Management and Logistics, Lund University, Faculty of Engineering, 2013, ISBN 978-91-7473-509-3, Media-Tryck, Lund University, Lund.

Howard C., J. Marklund. 2011. Evaluation of stock allocation policies in a divergent system with shipment consolidation. *European Journal of Operational Research* **211** 289–309.

Howard, C., O. Stenius 2013. Partial or Complete Deliveries in Two-echelon Inventory Systems?, in *Howard, C., Real-time Allocation Decisions in Multi-echelon Inventory Control*, Doctoral Thesis, Div. of Production Management, Dept. of Industrial Management and Logistics, Lund University, Faculty of Engineering, 2013, ISBN 978-91-7473-509-3, Media-Tryck, Lund University, Lund.

Jackson, P.L. 1988. Stock Allocation in a Two-Echelon Distribution System or What to do Until Your Ship Comes In. *Management Science* **34** 880–895.

Kapuscinski, R., Zhang, R.Q., Carbonneau, P., Moore, R. Reeves, B., 2004. Inventory Decisions in Dell's Supply Chain. *Interfaces*, **34**, 191–205.

Kiesmüller, G.P., A.G. de Kok, S.R. Smits, P.J.M. van Laarhoven. 2004. Evaluation of divergent N-echelon (s,nQ)-policies under compound renewal demand. *OR Spectrum* **26** 547–577.

Kuhn, H., M.G. Sternbeck, 2013. Integrative retail logistics: An exploratory study. *Operations Management Research* **6** 2–18.

Lengu D., A.A. Syntetos, , M.Z. Babai. 2014. Spare parts management: Linking distributional assumptions to demand classification. *European Journal of Operational Research*, **235**, 624–635.

Marklund J. 2002. Centralized Inventory Control in a Two-Level Distribution System with Poisson Demand. *Naval Research Logistics* **49** 798–822.

Marklund J. 2006. Controlling Inventories in Divergent Supply Chains with Advance-Order Information. *Operations Research* **54** 988–1010.

Marklund, J. 2011. Inventory Control in Divergent Supply Chains with Time-Based Dispatching and Shipment Consolidation. *Naval Research Logistics* **58** 59–71.

Marklund, J., K. Rosling. 2012. Lower bounds and heuristics for supply chain stock allocation, *Operations Research* **60** 92–105.

Marquès, G., C. Thierry, J. Lamothe, D. Gourc. 2010. A review of Vendor Managed Inventory (VMI): from concept to processes. *Production Planning and Control* **21** 547–561.

Moinzadeh K. 2002. A Multi-Echelon Inventory System with Information Exchange, *Management Science* **48** 414–426.

Mutlu, F., S. Çetinkaya, J.H. Bookbinder, 2010. An analytical model for computing the optimal time-and-quantity based policy for consolidated shipments. *IIE Transactions* **42** 367–377.

Özer, Ö. 2003. Replenishment strategies for distribution systems under advanced demand information. *Management Science* **49** 255–272.

Shang, K., Z. Tao, S. Zhou. 2015. Optimizing reorder intervals for two-echelon distribution systems with stochastic demand. *Operations Research* **63** 458–475.

Sherbrooke, C.C. 1968. METRIC: A Multi-Echelon Technique for Recoverable Item Control. *Operations Research* **16** 122–141.

Sherbrooke, C.C. 1986. Vari-Metric: Improved Approximations for Multi-Indenture, Multi-Echelon Availability Models, *Operations Research* **34** 311–319.

Simon, R.M. 1971. Stationary Properties of a Two Echelon Inventory Model for Low Demand Items. *Operations Research* **19** 761–777.

Zheng, Y.S. 1992. On properties of stochastic inventory systems. *Management Science* **38** 87–103.

Zipkin P. 2005. Foundations of Inventory Management. McGraw-Hill Primis.

E-Companion – Exact Analysis of Divergent Inventory Systems with Time-Based Shipment Consolidation and Compound Poisson Demand

Olof Stenius

Ayşe Gönül Karaarslan

Johan Marklund

A. G. de Kok

Appendix A: Algorithm for Determining TC_k^l

To find the lower bound for the total costs directly associated with retailer group k, TC_k, in Lemma 5,

$$TC_k^l = \min_{T_k, S_i} (TC_k | B_i(t_0) = 0, \forall \, i), \tag{A1}$$

we go through the following steps:

1. Determine a near optimal shipment interval for retailer group k, $\overline{T}_k$, using (29)

2. Given $\overline{T}_k$, determine near optimal reorder points for each retailer $\overline{S}_i$. This is done by optimizing each retailer separately as a single-echelon system. Note that the convexity property in Proposition 3 holds also for systems where $B_i = 0$ ($R_0 \to \infty$).

3. Calculate $\overline{TC}_k = TC_k(\overline{T}_k, \overline{S}_i)$.

4. Obtain upper and lower bounds for T_k for the optimization. These bounds are obtained analogously to the bounds in Proposition 5 as

$$T_k^l = \frac{\overline{TC}_k - \sqrt{\overline{TC}_k^2 - 2\omega_k \sum_{i \in N_k} h_0 \lambda_i \mu_i}}{\sum_{i \in N_k} h_0 \lambda_i \mu_i} \tag{A2}$$

$$T_k^u = \frac{\overline{TC}_k + \sqrt{\overline{TC}_k^2 - 2\omega_k \sum_{i \in N_k} h_0 \lambda_i \mu_i}}{\sum_{i \in N_k} h_0 \lambda_i \mu_i}.. \tag{A3}$$

5. Search through all values of T_k in this interval and optimize S_i for each T_k to find the lower bound, TC_k^l.

Appendix B: Proofs

Proof of Proposition 1

Proof. The reserved stock on hand at the warehouse increase whenever an unreserved unit on hand is reserved or a backordered unit arrives to the central warehouse. Thus the process by which the reserved stock on hand accumulates depends on the customer demand process, and the warehouse replenishment process. For fixed policies these two processes are in the current system independent of the shipment process, i.e., the length of the shipment interval, and when shipments leave the central warehouse. Hence, the same holds for the accumulation process of the reserved stock on hand. Moreover, as all unsatisfied demand is backordered, and all units will be reserved stock on hand at the warehouse at some point in time, the average rate by which the reserved stock on hand for retailer i accumulate is equal to the demand rate at retailer i, $\lambda_i \mu_i$. Consequently, if the previous shipment to retailer i left at time t, the expected number of units on hand at the warehouse reserved by retailer i, at $t + \tau$, for any $\tau \in (0, T_i]$, is $\lambda_i \mu_i \tau$. This means that the expected amount of reserved stock on hand for retailer i will increase linearly between two consecutive shipments from 0 to $\lambda_i \mu_i T_i$. Taking the average over time renders (4). $\square$

Proof of Lemma 1:

Proof. Conditioning on at least n retailer orders during $(t_0 - L_0, t_0]$, $\Phi_0 \geq n$, and that $\Psi^{n-1} = x$, for $x > 0$, there will be $u > 0$ partial backorders for retailer i caused by the nth retailer order if two conditions are fulfilled: (i) The nth retailer order originates from retailer i. The probability for this is $\frac{\lambda_i}{\lambda_0}$. (ii) The quantity of this order is $x + u$. The probability for this is $P\{Y_i = x + u\}$. This renders for $u > 0$,

$$P\{\hat{B}_i^n(t_0) = u | \Phi_0 \geq n \text{ and } \Psi^{n-1} = x\} = \frac{\lambda_i}{\lambda_0} P\{Y_i = x + u\}, \forall \, x > 0. \tag{B1}$$

(14) follows from unconditioning with respect to Φ_0 and Ψ^{n-1}. The latter by considering all possible positive values $x \in [1, S_0 - n + 1]$.

In order to have 0 partial backorders for retailer i, when the nth retailer order after $t_0 - L_0$ brings the nominal inventory position to a non-positive value, the proof is analogous. In this case, however, there are two different scenarios; either the nth customer arrives from retailer i and demands exactly x units to move the nominal inventory position to 0, or the nth customer arrives from another retailer and demands more than or equal to x units. $\square$

Proof of Lemma 2:

Proof. The distribution of complete backorders depends on n, the retailer order that brings the nominal inventory position to a non-positive value. However, because of the memoryless property of the compound Poisson demand, the complete backorders are independent of which retailers these n first orders originated from, and the sizes of these orders.

Given that there are a orders to retailer i after the nth retailer order, i.e., $\check{\Phi}_{i,n} = a$, it is clear that $P\{\check{B}_{i,n}(t_0) = v | \check{\Phi}_{i,n} = a\} = P\{Y_i^a = v\}$. (17) follows by taking the expectation over all possible outcomes of $\check{\Phi}_{i,n}$.

To arrive at (18) we note that given m retailer orders that are completely backordered, the probability that a of these originates with retailer i is $P\{Z_i^m = a\}$. Moreover, the probability of $n + m$ retailer orders in $(t_0 - L_0, t_0]$ given at least n orders in $(t_0 - L_0, t_0]$ is $P\{\Phi_0 = n + m\}/P\{\Phi_0 \geq n\}$. (18) follows as an expectation over all possible values of m. $\square$

Proof of Proposition 2:

Proof. By definition $P\{\hat{B}_i^n(t_0) = u\}$ is the probability that the nth retailer order brings Ψ^n to a non-positive value, causing u partial backorders for retailer i. Also, by definition $P\{\check{B}_{i,n}(t_0) = r - u\}$ is the probability that there are $r - u$ complete backorders for retailer i conditioned on that the nth retailer order brings Ψ^n to a non-positive value. Taking the expectation over all possible values of n and u (noting that $u \leq r$) renders (20).

For $r = 0$ we also need to consider the probability that the inventory position never reaches zero during the replenishment lead time, p_0, rendering (21). $\square$

Proof of Lemma 3:

Proof. In order for $\tilde{B}_i^{-m}(t_0) = u$ there are two possible scenarios; (a) The mth order before $(t_0 - L_0)$ arrives from retailer $j \neq i$ and the size of this order is at least $-\Psi^{-(m-1)} = -x$ units $(x < 0)$, and (b) the mth order originates with retailer i and is for at least $-\Psi^{-m} = -x$ units. Starting with (a) the probability that the mth customer arrives from retailer $j \neq i$ and demands more than $-x$ units is $(\lambda_{i^c}/\lambda_0)P\{Y_{i^c} \geq -x\}$. In order for $\tilde{B}_i^{-m}(t_0) = u$ in this scenario, the $m-1$ next customer orders need to contain u units to retailer i and need to assure that the nominal inventory position is x, which can be expressed as the probability $P\{V_i^{m-1} = u \text{ and } \Psi^{-(m-1)} = x\}$. A summation over all possible values of x $(x \in [S_0 + m - 1, -1])$ generates the first part of (25).

The probability for scenario (b) is $(\lambda_i/\lambda_0)P\{Y_i \geq -x\}$. In this scenario, $-x$ units of the mth order will be backordered at t_0. Thus, in order for $\tilde{B}_i^{-m}(t_0) = u$, the next $m-1$ customers need to order $u - (-x) = u + x$ units to retailer i and ensure that $\Psi^{-(m-1)} = x$, which can be expressed as the probability $P\{V_i^{m-1} = u + x \text{ and } \Psi^{-(m-1)} = x\}$. A summation over all possible values of x renders the second part of (25).

In order for $V_i^{m-1} = u$ and $\Psi^{-(m-1)} = x$, the last $m-1$ customer orders before $t_0 - L_0$ need to include u units to retailer i and $x - S_0 - u$ units to all other retailers (recall $S_0 < 0$ and $x < 0$). With Z_i^{m-1} defined as in Section 3.3.1 and determined by (16) we get (26). $\qquad\square$

Proof of Proposition 3:

Proof. Neither the shipment costs, $\sum_{k=1}^{K} \frac{\omega_k}{T_k}$, nor the holding costs at the central warehouse are affected by S_i. Furthermore, the holding cost and backorder cost at retailer i are unaffected by the order-up-to levels at other retailers. Thus, for fixed R_0 and $\mathbf{T}$, $TC(R_0, \mathbf{S}, \mathbf{T})$ is separable in the retailer order-up-to levels. To assert the convexity in S_i we define the holding and backorder costs at retailer i as

$$RC_i(R_0, S_i, T_i) = h_i E[IL_i^+(s)] + \beta_i E[IL_i^-(s)]. \tag{B2}$$

It is sufficient to show convexity for $RC_i(R_0, S_i, T_i)$ with respect to S_i for each retailer i. We define the difference function $\Delta G(s)$ as follows:

$$\Delta G(s) = RC_i(R_0, s + 1, T_i) - RC_i(R_0, s, T_i). \tag{B3}$$

To prove convexity, we need to show that $\Delta G(s) - \Delta G(s-1) \geq 0$. First, by using (7), (8) and (9), we rewrite $RC_i(R_0, s, T_i)$ as:

$$
\begin{aligned}
RC_i(R_0, s, T_i) &= (h_i + \beta_i)E[IL_i^+(s)] - \beta_i E[IL_i(s)], \\
&= (h_i + \beta_i)\frac{1}{T_i}\sum_{j=1}^{s}\sum_{r=0}^{s-j} j P\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s - j - r\}dx, \\
&\quad -\beta_i(s - E[B_i(t_0)] - \lambda_i \mu_i L_i).
\end{aligned} \tag{B4}
$$

Note that the probability mass function and expectation of $B_i(t_0)$ does not depend on the order-up-to

levels of the retailers. Next, we derive $RC_i(R_0, s+1, T_i)$ in terms of $RC_i(R_0, s, T_i)$:

$$RC_i(R_0, s+1, T_i) =$$

$$(h_i + \beta_i)\frac{1}{T_i}\sum_{j=1}^{s+1}\sum_{r=0}^{s+1-j} jP\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s+1-j-r\}dx$$

$$-\beta_i(s+1 - E[B_i(t_0)] - \lambda_i\mu_i L_i),$$

$$= (h_i + \beta_i)\frac{1}{T_i}\sum_{z=0}^{s}\sum_{r=0}^{s-z}(z+1)P\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s-z-r\}dx$$

$$-\beta_i(s+1 - E[B_i(t_0)] - \lambda_i\mu_i L_i),$$

$$= (h_i + \beta_i)\frac{1}{T_i}\sum_{z=0}^{s}\sum_{r=0}^{s-z}zP\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s-z-r\}dx$$

$$+(h_i + \beta_i)\frac{1}{T_i}\sum_{z=0}^{s}\sum_{r=0}^{s-z}P\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s-z-r\}dx$$

$$-\beta_i(s - E[B_i(t_0)] - \lambda_i\mu_i L_i) - \beta_i,$$

$$= (h_i + \beta_i)\frac{1}{T_i}\sum_{z=0}^{s}\sum_{r=0}^{s-z}P\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s-z-r\}dx$$

$$-\beta_i + RC_i(R_0, s, T_i). \tag{B5}$$

As a result the first order difference is equal to:

$$\Delta G(s) = (h_i + \beta_i)\frac{1}{T_i}\sum_{j=0}^{s}\sum_{r=0}^{s-j}P\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s-j-r\}dx - \beta_i. \tag{B6}$$

By following the similar line of thought, we get the following for the second order difference:

$$\Delta G(s) - \Delta G(s-1) = (h_i + \beta_i)\frac{1}{T_i}\sum_{r=0}^{s}P\{B_i(t_0) = r\}\int_0^{T_i} P\{D_i(L_i + x) = s-r\}dx \geq 0. \tag{B7}$$

$\square$

Proof of Proposition 4:

Proof. It is clear from (3) that $h_0 E[IL_0^+]$ is increasing in R_0 for all values of $R_0 \geq -Q_0$ and that the total cost $TC(R_0, \mathbf{S}, \mathbf{T}) = h_0 E[IL_0^+] + \sum_{k \in K} TC_k(R_0, \mathbf{S}, T_k) \geq h_0 E[IL_0^+] + \sum_{k \in K} TC_k^l$. With $R_0^u = min\{R_0 : h_0 E[IL_0^+] + \sum_{k \in K} TC_k^l \geq \overline{TC}\}$ it follows that for all $R_0 > R_0^u$, $TC(R_0, \mathbf{S}, \mathbf{T}) \geq \overline{TC}$ and searching this region cannot reduce the costs. $\square$

Proof of Proposition 5:

Proof. Let

$$\Theta_k(T_k) = h_0 \sum_{i \in N_k} E[W_i] + \frac{\omega_k}{T_k} = \frac{1}{2} \sum_{i \in N_k} h_0 \lambda_i \mu_i T_k + \frac{\omega_k}{T_k}, \tag{B8}$$

be the costs directly related to shipment interval T_k. Note that $\Theta_k(T_k)$ is strictly convex in T_k as (for $T_k > 0$)

$$\frac{\partial^2 \Theta_k}{\partial T_k^2} = \frac{\partial^2 (\frac{1}{2} \sum_{i \in N_k} h_0 \lambda_i \mu_i T_k + \frac{\omega_k}{T_k})}{\partial T_k^2} = 0 + 2 \frac{\omega_k}{T_k^3} > 0. \tag{B9}$$

Note also that

$$TC(R_0, \mathbf{S}, \mathbf{T}) = h_0 E[IL_0^+] + \sum_{\kappa \in K} TC_\kappa(R_0, \mathbf{S}, T_\kappa)$$

$$\geq TC_k(R_0, \mathbf{S}, T_k) + \sum_{\kappa \in K \setminus \{k\}} TC_\kappa(R_0, \mathbf{S}, T_\kappa) \geq \Theta_k(T_k) + \sum_{\kappa \in K \setminus \{k\}} TC_\kappa^l. \tag{B10}$$

It follows that no T_k satisfying

$$\Theta_k(T_k) \geq \overline{TC} - \sum_{\kappa \in K \setminus \{k\}} TC_\kappa^l \tag{B11}$$

can render a lower total expected cost than $\overline{TC}$. From (B8), (B11) and the convexity of $\Theta_k(T_k)$ we arrive at (31) and (32) by solving

$$\frac{1}{2} \sum_{i \in N_k} h_0 \lambda_i \mu_i T_k + \frac{\omega_k}{T_k} = \overline{TC} - \sum_{\kappa \in K \setminus \{k\}} TC_\kappa^l \tag{B12}$$

with respect to T_k. When T_k is smaller than the smallest root, T_k^l, or larger than the largest root, T_k^u, (B11) is always satisfied and these regions do not need to be searched. □

Proof of Lemma 4:

Proof. From Proposition 5 we know that the costs directly related to the shipment interval T_k, $\Theta_k(T_k) = \sum_{i \in N_k} E[W_i] + \frac{\omega_k}{T_k}$ is convex in T_k. Hence, the shipment interval $T_{k,l}$ that minimizes $\Theta_k(T_k)$ is obtained from the first order optimality condition

$$\frac{\delta \Theta_k}{\delta T_k} = \frac{\delta}{\delta T_k} \left(\frac{1}{2} \sum_{i \in N_k} h_0 \lambda_i \mu_i T_k + \frac{\omega_k}{T_k} \right) = 0, \tag{B13}$$

which renders

$$T_{k,l} = \sqrt{\frac{2\omega_k}{\sum_{i \in N_k} h_0 \lambda_i \mu_i}}. \tag{B14}$$

(33) follows from

$$TC_k(T_k) \geq \Theta_k(T_k) \geq \Theta_k(T_{k,l}) = \sqrt{2 h_0 \omega_k \sum_{i \in N_k} \lambda_i \mu_i}. \tag{B15}$$

□

Proof of Lemma 5:

Proof. From (27) $TC_k(R_0, \mathbf{S}, T_k) = \frac{\omega_k}{T_k} + \sum_{i \in N_k} h_0 E[W_i] + h_i E[IL_i^+] + \beta_i E[IL_i^-]$. Note that neither W_i nor $\frac{\omega_k}{T_k}$ depend on R_0. Moreover, the retailer costs, $RC_i(R_0, S_i, T_k) = h_i E[IL_i^+(s)] + \beta_i E[IL_i^-(s)]$, depend on R_0 solely through the backorder distribution, $B_i(t_0)$. From (6) we can show that the probability of an inventory level j at time $t_0 + L_i + t$ at retailer i when $B_i(t_0) = b_i$ is

$$
\begin{aligned}
P\{IL_i(t_0 + L_i + t) = j\} &= P\{b_i + D_i(t_0, t_0 + L_i + t) = S_i - j\} \\
&= P\{(b_i + 1) + D_i(t_0, t_0 + L_i + t) = (S_i + 1) - j\}, \quad \text{(B16)}
\end{aligned}
$$

which implies that

$$
RC_i(S_i, T_k | B_i(t_0) = b_i) = RC_i(S_i + 1, T_k | B_i(t_0) = b_i + 1). \quad \text{(B17)}
$$

For reasons of exposition and without loss of generality we renumber the retailers in retailer group k, $\{1, 2, ..., N_k\}$. Now, let $\mathbf{S}_{N_k}$ denote the vector of all order-up-to levels within retailer group k, $\{S_1, ..., S_{N_k}\}$, and $\mathbf{B}_{N_k}(t_0)$ denote the vector of backordered units to each retailer in retailer group k at t_0, $\{B_1(t_0), ..., B_{N_k}(t_0)\}$. The total costs for retailer group k, $TC_k(R_0, \mathbf{S}_{N_k}, T_k)$, for any values of R_0, $\mathbf{S}_{N_k}$ and T_k can then be seen as a sum over weighted averages of costs for all possible backorder combinations:

$$
TC_k(R_0, \mathbf{S}_{N_k}, T_k) = \sum_{b_1=0}^{\infty} ... \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, ..., b_k\}\} TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{b_1, ..., b_k\}). \quad \text{(B18)}
$$

This gives us for any value of R_0

$$
\begin{aligned}
TC_k(R_0, \mathbf{S}, T_k) &\geq \min_{T_k, \mathbf{S}_{N_k}} [TC_k(R_0, \mathbf{S}, T_k)] \\
&= \min_{T_k, \mathbf{S}_{N_k}} \left[\sum_{b_1=0}^{\infty} ... \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, ..., b_k\}\} TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{b_1, ..., b_k\}) \right] \\
&\geq \sum_{b_1=0}^{\infty} ... \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, ..., b_k\}\} \min_{T_k, \mathbf{S}_{N_k}} [TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{b_1, ..., b_k\})] \\
&= \sum_{b_1=0}^{\infty} ... \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, ..., b_k\}\} \min_{T_k, \mathbf{S}_{N_k}} [TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{0, ..., 0\})] \\
&= \min_{T_k, S_i \forall\ i \in N_k} (TC_k | B_i(t_0) = 0, \forall\ i \in N_k). \quad \text{(B19)}
\end{aligned}
$$

The first equality follows from (B18). The second inequality is a consequence of relaxing the constraint forcing the same values of T_k and $\mathbf{S}_{N_k}$ for all values of $\mathbf{B}_{N_k}(t_0)$. The second equality follows from (B17) and the fact that only the retailer costs are affected by the backorder distribution. The last equality follows directly as probabilities must sum to 1, completing the proof of (34) □

Appendix C: A Small Numerical Example

To illustrate the analysis we consider a system consisting of 3 retailers belonging to 2 retailer groups; retailers $\{1, 2\}$ constitute the first group (and therefore have equal shipment intervals) and retailer 3 the second. Each retailer face compound Poisson demand with a logarithmic compounding distribution, i.e. with a variance to mean ratio of the demand at retailer i of $\rho_i = Var[D_i]/E[D_i]$ we have: $P\{Y_i = y\} = -\alpha_i^y/(ln(1 - \alpha_i)y)$ and $\lambda_i = -E[D_i](1 - \alpha_i)ln(1 - \alpha_i)/\alpha_i$, where $\alpha_i = 1 - \rho_i^{-1}$. The considered problem parameters are presented in Table 2.

Table 2: Values for parameters and decision variables in the illustrative example

$E[D_i] = \{1, 1, 1\}$	$R_0 = -2$	$T_i = \{0.5, 0.5, 1\}$	$L_i = \{0.5, 0.5, 1, 0.5\}$	$\beta_i = \{10, 10, 10\}$
$\rho_i = \{4, 2, 1.5\}$	$Q_0 = 5$	$S_i = \{4, 4, 4\}$	$h_i = \{1, 1, 1, 1\}$	$\omega_k = \{2, 2\}$

As seen in Section 3.2, to analyze retailers i $(i = 1, 2, 3)$ we need to determine the probability of r warehouse backorders designated to retailer i, $P\{B_i(t_0) = r\}$, for all $r = [0, S_i - 1]$, and the expected amount of backorders to retailer i, $E[B_i(t_0)]$. The computations are based on the analysis in Section 3.3. Two examples of the backorder distribution to retailer 1 conditioned on the inventory positions are $\{P\{B_1(t_0) = r|IP_0(t_0 - L_0) = -1\}, r = 0, 1, 2, 3\} = \{0.607, 0.262, 0.057, 0.028\}$ and $\{P\{B_1(t_0) = r|IP_0(t_0 - L_0) = 3\}, r = 0, 1, 2, 3\} = \{0.942, 0.023, 0.012, 0.008\}$. Taking the average over all possible inventory positions $(IP_0(t_0 - L_0) = [-1, 3])$ we get the steady state distributions of the backorders designated to each retailer, presented in Table 3. Exemplifying the Remark on page 20, Table 3 also presents the expected backorders designated to each retailer and illustrates their disproportions to the demand per time unit. One result of this is that the delay in the expected replenishment lead time to retailer 1, caused by backorders at the central warehouse, will be higher than the delay experienced by retailers 2 and 3.

Table 3: Distribution of warehouse backorders designated to each retailer

Retailer (i)	$\{P\{B_i(t_0) = r\}, r = 0, 1, 2, 3\}$	$E[B_i(t_0)]$
1	$\{0.824, 0.096, 0.032, 0.017\}$	0.399
2	$\{0.773, 0.144, 0.048, 0.020\}$	0.373
3	$\{0.754, 0.165, 0.054, 0.018\}$	0.367

Knowing the backorder distributions, the expected stock on hand, $E[IL_i^+]$, the expected backorders, $E[IL_i^-]$, and the fill rates, γ_i, at the retailers can be determined from (7), (9) and (10) respectively. Calculating the expected stock on hand at the central warehouse using (3) and (4), the total cost of the system can be determined from (2). The results, determined analytically by the suggested approach, and simulated in a discrete event simulation program (Extend), are presented in Table 4.

Table 4: Results from exact analysis and simulation (Sim)

	$E[IL_0^+]$	$E[IL_i^+]$	$E[IL_i^-]$	TC	$\gamma_i(\%)$
Exact	1.639	$\{3.087, 2.541, 2.704\}$	$\{0.236, 0.165, 0.071\}$	20.691	$\{72.6, 79.5, 88.1\}$
Sim[1]	1.639	$\{3.087, 2.541, 2.704\}$	$\{0.236, 0.165, 0.071\}$	20.691	$\{72.6, 79.5, 88.1\}$

[1] The Standard deviations of the simulated results were < 0.001.

Appendix D: Numerical Study

For all 128 problem settings defined in Section 5, the shipment intervals, $T_k \; \forall \; k \in K$, the reorder points at the central warehouse, R_0, and the order up to levels at the retailers, $S_i \; \forall \; i \in N$, are optimized using the method described in Section 4. For optimizing the shipment intervals we have used a step size of of 0.01 time units. The complete results for all settings are available from the authors upon request. Table 5 summarizes the results in terms of average effects on: the optimal total cost, TC^*, the optimal reorder point at the central warehouse, R_0^*, the average of the optimal order-up-to level at the retailers, S_i^*, and the average of the optimal shipment intervals, T_k^*. It also includes the relative difference between the heuristic shipment intervals and the optimal, ΔT, the relative increase in the associated costs, ΔC, the difference between the upper bound on R_0, R_0^u, and the optimal value, R_0^*, ΔR_0^u, the difference between the optimal value on T_k, T_k^*, and the lower bound, T_k^l, ΔT_k^l, and the difference between the upper bound on T_k, T_k^u, and the optimal value, T_k^*, ΔT_k. The results associated with $\rho_i = 1$ are averages across all 64 problems where $\rho_i = 1$, and analogously for all other parameters.

Table 5: Average results for the test series, for low and high values of N, ρ_i, β_i, ω_k, L_0, L_i and Q_0 as well as averages over all problems and minimum and maximum values

	$E[TC^*]$	$E[R_0^*]$	$E[S_i^*]$	$E[T_k^*]$	$E[\Delta C]$	$E[\Delta T]$	$E[\Delta R_0^u]$	$E[\Delta T_k^l]$	$E[\Delta T_k^u]$
$\rho_i{}^1 = 1$	82.02	19.11	15.08	3.24	0.12%	-0.60%	4.91	2.05	15.81
$\rho_i{}^1 = 5$	141.00	20.67	23.75	3.05	0.16%	6.57%	9.92	2.33	31.28
$\beta_i = 10$	93.66	18.19	15.88	3.30	0.08%	-2.27%	7.42	2.21	18.88
$\beta_i = 100$	129.37	21.59	22.94	2.99	0.21%	8.24%	7.41	2.17	28.21
$\omega_k = 10$	81.96	21.25	15.97	1.54	0.13%	1.90%	7.28	1.25	19.07
$\omega_k = 100$	141.07	18.53	22.85	4.75	0.15%	4.07%	7.55	3.13	28.02
$N = 3$	84.78	10.86	20.45	3.68	0.15%	3.18%	6.53	2.52	25.82
$N = 6$	138.24	28.92	18.38	2.61	0.14%	2.79%	8.30	1.86	21.27
$L_0 = 1$	108.36	1.27	18.86	3.11	0.17%	4.41%	5.06	2.11	22.05
$L_0 = 5$	114.66	38.52	19.96	3.18	0.11%	1.56%	9.77	2.27	25.04
$L_i = 1$ and 2	107.98	20.45	17.26	3.10	0.18%	4.94%	7.48	2.13	22.96
$L_i = 2$ and 4	115.04	19.33	21.57	3.19	0.11%	1.03%	7.34	2.26	24.14
$Q_0 = 2$	110.40	24.50	19.13	3.13	0.15%	3.49%	6.33	2.15	22.96
$Q_0 = 20$	112.63	15.28	19.69	3.16	0.14%	2.48%	8.50	2.23	24.13
Average	111.51	19.89	19.41	3.15	0.14%	2.99%	7.41	2.19	23.55
Minimum	31.61	-10	7.67	1.09	0.00%	-9.79%	0	0.91	7.05
Maximum	264.19	68	38.33	6.11	0.66%	18.60%	24	4.23	51.26

[1]$\rho_i = 1$ corresponds to Poisson demand at all retailers and $\rho_i = 5$ to compound Poisson demand with logarithmic compounding distributions so that the variance-to-mean ratio of the demand per time unit at retailer i is $5 \; \forall \; i \in N$.

Focusing first on the computational aspects and the optimality bounds, it is relevant to know that the optimization times for studied problems were between 0.2 and 120 minutes on a Dell Latitude 6400 lap top. The parameters that seem to affect the computational times the most are λ_0, L_0, N and K. An important observation is that the fairly time consuming calculations of the backorder distributions at the central warehouse only needs to be computed once for each value of R_0 (in a given problem). Especially for the computationally more demanding problems, most of the computational time was spent on calculating these distributions. An explanation for this is that the complexity of determining these distributions increase with $R_0 + Q_0$. As a result, the upper bound on the reorder level at the central warehouse, R_0^u is the most important bound. Table 5 shows that this bound is rather tight,

exceeding the optimal value R_0^* by on average with only 7.41 units. It is especially tight in the cases where $\rho_i = 1$ (on average only 4.91 units above optimum) and when $L_0 = 1$ (on average only 5.06 units above optimum). In fact, in the 32 problem settings investigated, where $\rho_i = 1$ and $L_0 = 1$, $E[\Delta R_0^u] = 3.53$. The fact that $E[\Delta R_0^u]$ is lower in systems where L_0 is lower is intuitive, as the optimal reorder level, R_0^*, is much lower for these systems. The effect of ρ_i on the performance of the bound is perhaps less obvious. In order to explain this, recall that this bound is based on an estimation of the minimum costs at the retailers, TC_k^l, that assumes no backorders at the central warehouse (see Proposition 4). When the demand has a higher variance-to-mean ratio, a higher R_0-value is required for reaching this situation and the bound will therefore become looser.

Turning to the bounds on T_k, Table 5 shows that the lower bound is on average 2.19 time units below the optimal value (which is on average 3.15). The upper bound is looser as it is on average 23.55 time units above the optimal value. These bounds play a less important role in reducing the computational time than the upper bound on R_0^*.

Shifting our attention to the EOI heuristic, Marklund (2011) shows that it performs very well for Poisson demand. Based on Table 5 this seems to hold also for compound Poisson demand. The expected relative cost increase for all problem settings is only 0.14%, although the heuristic tends to overestimate the optimal shipment intervals with on average 2.99%. The relative cost increase is also only slightly higher in the systems where the variability in the demand is high 0.16% compared to 0.12% for the Poisson systems. There is a stronger tendency to overestimate the shipment intervals for the systems where $\rho_i = 5$ ($E[\Delta T] = 6.57\%$), but because the total costs are much higher in these systems, the relative increase is still small. The parameter that seems to have the biggest influence on the performance of the heuristic is, in fact, the backorder costs. In systems where the backorder costs are high $E[\Delta C] = 0.21\%$. An explanation may be the desire to increase the flexibility in these systems by reducing the shipment intervals (the shipment intervals are overestimated by 8.24% in systems where $\beta_i = 100$ when using the heuristic).

Considering the behavior of the optimal solutions, Table 5 illustrates that when the variability increases (i.e. comparing $\rho_i = 1$ to $\rho_i = 5$), the biggest difference in the control parameters is seen in the order-up-to levels of the retailers, which increase from on average 15.08 to 23.75. However there are also effects on the shipment intervals and reorder points at the central warehouse. The optimal warehouse reorder point increases from on average 19.11 to 20.67, thus raising the amount of available units at the central warehouse to handle the variability. The shipment intervals decrease from on average 3.24 to 3.05 with the effect that there is more flexibility in the system (the system can react faster if there is a big order at a specific retailer). Similar effects can be seen when increasing the backorder cost (β_i). The largest effect on the the optimal control parameters is an increase in the average order-up-to levels at the retailers from 15.88 to 22.94. However, we also see an increase in the average reorder point at the central warehouse from 18.19 to 21.59 and a decrease in the average shipment intervals (from 3.30 to 2.99), increasing the flexibility.

With regards to the shipment costs, we can see from Table 5 that as $\omega_k \ \forall \ k$ increase, the system reacts by increasing the shipment intervals (from on average 1.54 to 4.75). Moreover, the order-up-to levels at the retailers need to be raised accordingly in order to ensure stock for a longer replenishment cycle. Maybe less intuitively, the average optimal reorder points at the central warehouse decreases for these systems. This can be explained by the fact that increased shipment intervals results in longer replenishment lead times to the retailers, which may reduce the relative impact of inventory pooling at the central warehouse. Another contributing factor may be that the consolidation stock at the central warehouse increases with the shipment intervals. Thus, in order to avoid too much stock at the warehouse, the reorder point is reduced. For the other parameters, N, L_0, L_i and Q_0, the behavior of the optimal solutions did not offer any insights beyond the obvious.

Paper III

Partial or Complete Deliveries in Two-echelon Inventory Systems?

Christian Howard ● Olof Stenius

Department of Industrial Management and Logistics, Lund University
christian.howard@iml.lth.se ● olle.stenius@iml.lth.se

Abstract

We consider a continuous review inventory model consisting of a central warehouse supplying N retailers which face stochastic demand. All installations replenish using reorder point policies with fixed batch sizes. The focus is on evaluating different central warehouse delivery polices. If the central warehouse cannot satisfy an entire order immediately, previous models predominantly assume that any available units are shipped immediately (partial delivery). However, depending on the cost structure and the current state of the system it may be more effective to wait until the entire order is available (complete delivery). We introduce a new state-dependent delivery policy where a cost minimizing decision between partial or complete deliveries is made for each occurring order. We provide an exact method for cost evaluation and optimization of the reorder points under this policy, as well as for the pure partial - and for the pure complete delivery policies. We also derive sufficient conditions for when complete deliveries should always be chosen over partial deliveries. Numerical results show that significant benefits can be reaped by using our new policy.

1. Introduction

Increasing fuel prices, tighter environmental legislation and the growing strive to create sustainable supply chains draws focus to the distribution and transportation aspects of inventory control. As the importance of these issues increases, so does the importance of designing good policies for when and how to ship physical products within a supply chain. In this work we consider a two-echelon continuous review inventory system with a central warehouse and a number of retailers facing stochastic Poisson demand. All installations order in batches using reorder point policies (commonly referred to as installation stock (R,Q) policies). In previous exact analysis of such systems it is generally assumed that if shortages at the central warehouse occur, any units currently available are shipped to the retailers as soon as possible (Axsäter, 2003). This is referred to as a partial delivery policy and it means that a given order may be shipped to the retailer in several parts in varying sizes and at different times. An obvious alternative is the complete delivery policy, where units are always shipped in complete batches. A question not analyzed in previous models is how the choice of delivery policy may affect the operating characteristics of the system studied. This is particularly noticeable in models featuring partial deliveries as it is assumed that no extra cost is incurred, regardless of how many separate shipments are required to fulfill a given retailer order. This is in many cases a poor representation of reality, where splitting an order results in repeating activities such as order picking, loading, unloading, receiving, inspection, authorization and invoicing. Add to this the extra cost of transportation (for instance, being forced to dispatch several trucks) as well as the environmental consequences of such actions, and it is clear that the delivery strategy is an important part of controlling inventories efficiently.

In this work we present a new model that incorporates, and makes it possible to exactly evaluate, the impact of different delivery policies. In addition to standard holding - and backorder costs we consider what we refer to as an (extra) handling cost. This handling cost is a fixed charge for each partial delivery and it quantifies the extra cost of partial delivery compared to complete delivery of a given retailer order. Hence, the extra cost of activities such as the ones mentioned above, as well as quantifiable environmental costs can be included in this cost parameter. Given this new, more general, cost structure it is easy to spot the weaknesses in both the partial - and complete delivery policies. For example, assume that a retailer orders Q units and the central warehouse has only one unit currently available, with $Q-1$ more units arriving in stock within a very short time period. Partial delivery implies that one unit will be shipped immediately, and the remaining $Q-1$ will be shipped just moments later, incurring an unnecessary extra handling cost. Conversely, complete

delivery can lead to situations where a retailer desperately needs replenishment, justifying the extra handling cost, but no units are shipped because the order is not yet complete. In light of these predicaments, we introduce a new Mixed State-Dependent delivery policy, referred to as the MSD policy. Under the MSD policy, a decision between delivering a retailer order in one or several parts is made at the time of order placement. The decision minimizes the expected costs for the entire system and is based on the arrival times of incoming orders to the central warehouse. With the advances in information technology such information is becoming readily available. We provide an exact method for cost evaluation and optimization of all reorder points in the system for this new policy, as well as for the partial and complete delivery policies.

Looking at the literature, there are quite a few models, both exact and approximate, dealing with one warehouse - multiple retailer systems. For general overviews of this literature we refer to, for example, Axsäter (2003) and Marklund (2011). We will focus on models with exact solutions and non-identical retailers. Our emphasis on a state-dependent delivery policy that uses real-time information also means that we will focus on continuous review systems. For recent overviews of the periodic review literature see, for example, Chu and Shen (2010), Marklund and Rosling (2012) and references therein. Early contributions to the continuous review literature focus on one-for-one ordering (base-stock policies) and include Graves (1985) and Axsäter (1990). Assuming Poisson demand and First Come - First Served (FCFS) allocation, Graves (1985) provides the distribution of the retailer inventory levels using a technique based on binomial disaggregation of the central warehouse backorders. Through the distributions of the inventory levels it is easy to obtain performance measures such as the expected holding - and backorder costs. Axsäter (1990) uses a different technique to determine the expected holding - and backorder costs. The technique is based on tracking the costs that accrue as an arbitrary unit moves through the system. An extension of this technique to compound Poisson demand is available in Forsberg (1995).

Turning our attention to batch ordering policies, Axsäter (1993,1998) and Forsberg (1997) extend the unit tracking technique in Axsäter (1990) to installation stock (R,Q) policies. Axsäter (1997) and Chen and Zheng (1997) provide exact results for echelon stock (R,Q) policies. An important work is Axsäter (2000) which provides a model for installation stock policies, featuring (R,Q) policies and compound Poisson demand. Note that all papers mentioned so far assume partial delivery policies. If we consider the special case in our model where the handling cost is set to zero, we provide an alternative method of analysis for the models in Forsberg (1997) and Axsäter (2000) under Poisson demand. Andersson (1999) generalizes the technique in Axsäter (2000) and provides a cost evaluation method for the complete delivery policy. We provide an alternative cost evaluation

technique to this work as well when demand is Poisson. Noteworthy is that Andersson (1999) does not provide a method for optimizing the reorder points for the complete delivery policy. To the best of our knowledge, our work is the first to provide an optimization method for the case of complete deliveries.

Recently, there have been a number of papers considering new policies for the central warehouse. Marklund (2011) shares our focus on the central warehouse delivery policy, as it considers a time-based shipment consolidation policy, as well as an (R,Q) policy at the central warehouse and Poisson demand (later generalized to compound Poisson demand by Stenius et al., 2013). An important difference is that Marklund (2011) considers base-stock policies at the retailers, as opposed to our (R,Q) policies. Furthermore, Marklund (2011) assumes that shipments leave the warehouse at regular time intervals, whereas a state-dependent decision on when to release a batch (or part of a batch) is made with our MSD policy. Central warehouse policies also based on this type of extended information can be found in, for example, Marklund (2002), Marklund (2006), Axsäter and Marklund (2008). However, their focus is different from ours as they study the warehouse ordering policy, assuming partial deliveries. Axsäter and Marklund (2008) provide an optimal position-based ordering policy that relaxes the FCFS allocation assumption. Apart from this work, FCFS allocation is assumed in all papers mentioned above, and is also assumed in our current work. In addition to being a simple easily implemented allocation policy, there are a number of numerical studies suggesting that FCFS generally performs well for the type of system considered in this work (e.g., Graves 1996, Axsäter, 2007, Howard and Marklund, 2011, Howard, 2013).

The main contributions of this paper can be summarized as follows: First, we introduce a new state-dependent delivery policy and provide an exact method for cost analysis and optimization of the system reorder points, given this policy. Second, we generalize previous exact partial - and complete delivery models. For partial delivery models we introduce a new cost parameter. For complete delivery models we provide a method for optimizing the reorder points. Furthermore, in our analysis we use a different approach compared to the previous literature; one which we believe can be fruitfully applied to other problems. Finally, through analytical as well as numerical results we provide managerial insights on when partial or complete deliveries are reasonable to use, and the value of using a more advanced state-dependent delivery policy. For example, we derive sufficient conditions for when one should always choose complete deliveries. The numerical tests show that the new MSD policy can lead to significant cost reductions compared to the simpler policies. Over all problem scenarios considered in the study, the maximum expected total cost increase of using the partial delivery policy, compared to the MSD policy, was 26.6% (the average was 5.8%). The

equivalent maximum cost increase of the complete delivery policy was 17.9% (the average was 5.9%). Furthermore, the numerical tests suggest that, under our more general cost structure, it is optimal to keep more stock at the central warehouse compared to what has been reported in the previous literature.

2. Problem formulation

We consider a continuous review system consisting of one central warehouse and N non-identical retailers. The retailers face customer demand that occurs according to independent Poisson processes, and they place replenishment orders with the central warehouse. The central warehouse, in turn, places orders with an outside supplier with a constant lead time. All stock points apply complete backordering and demand is satisfied according to a First Come - First Served (FCFS) principle. Orders are placed using installation stock (R,Q) policies, where a batch of Q units is ordered when the inventory position (= inventory level + outstanding orders; inventory level = stock on hand – backorders) drops to or below R. As mentioned above, the lead time for a batch shipped from the outside supplier to the central warehouse is constant. The transportation times (including time for loading, shipping and receiving) from the central warehouse to the retailers are also constant. However, the lead times for orders placed by the retailers depend on the availability of units at the central warehouse, and the type of delivery policy used at this location. They are therefore stochastic. We assume linear holding costs per unit and time unit at all stock points and linear backorder costs per unit and time unit at the retailers. The holding costs for units in transport between the central warehouse and the retailers are not included in our model. This is because these costs are constant on average and independent of the choice of ordering – or delivery policy.

The delivery policy determines when units will be dispatched from the central warehouse. When one or more units have been dispatched, they will arrive at the retailer after the constant transportation time. We consider three different delivery policies: (i) the partial delivery policy, (ii) the complete delivery policy and (iii) the mixed state-dependent delivery policy. The partial delivery policy (referred to as the PD policy) means that any available units are dispatched from the central warehouse as soon as possible after they are ordered by a retailer. If this results in a batch of Q_i being delivered in multiple parts, we refer to this as a partial delivery of that batch. The complete delivery policy (CD policy) implies that no units are dispatched until the entire batch of Q_i is available at the central warehouse. In other words, complete delivery is applied to all incoming orders. Note that the PD and CD policy can result in the same course of action, whenever all Q_i units are available at the central warehouse at the same time. Applying the mixed state-dependent delivery policy (MSD

policy), a decision between partial or complete delivery is made at the time the retailer order occurs. That is, if part of (but not the entire) order will be available at some point in time, a choice is made between dispatching the first part at the earliest possible time or waiting until the entire batch is available. This choice is referred to as the MSD decision. The MSD decision is made according to what we call the MSD decision rule, which minimizes the expected costs for the entire system.

We assume that partial delivery can lead to a maximum of one extra delivery compared to complete delivery. That is, an order cannot be split in more than two parts. This will always be the case given some mild assumptions regarding the central warehouse and retailer batch sizes (this is discussed in further detail after the notation has been defined). There is a fixed handling cost for each batch that is partially delivered. This cost quantifies the extra costs that are incurred due to the two separate delivery occasions, compared to complete delivery where the entire batch of Q_i units is delivered on one single occasion.

The ordering policy at the central warehouse operates independently of the delivery policy. When applying complete delivery, there can be units physically at the central warehouse that have already been assigned to a specific retailer (recall that FCFS is used). These units, which are referred to as units on hold, are waiting for order completion before they can be dispatched. Units on hold are not included in the stock on hand (we define stock on hand to only include units that are available to satisfy future retailer orders). As a consequence, units on hold are not included in the inventory level or the inventory position (the state of which, triggers orders to the outside supplier). However, units on hold do incur holding costs at the central warehouse, just like the units on hand.

We introduce the following notation:

N	=	number of retailers
N	=	$\{1,2,…,N\}$, set of retailer indices
R_0	=	reorder point at the central warehouse
R_i	=	reorder point at retailer i
Q_0	=	order quantity at the central warehouse,
Q_i	=	order quantity at retailer i
q	=	largest common factor of $Q_0,Q_1,…,Q_N$
L_0	=	lead time for an order placed by the central warehouse with the outside supplier
L_i	=	transportation time for an order placed by retailer i with the central warehouse
h_0	=	holding cost per unit and time unit at the central warehouse, $h_0 > 0$
h_i	=	holding cost per unit and time unit at retailer i, $h_i > 0$
b_i	=	backorder cost per unit and time unit at retailer i, $b_i > 0$

θ_i	=	handling cost for each partially delivered batch to retailer i, $\theta_i \geq 0$
λ_i	=	average customer demand rate at retailer i
$IP_0(t)$	=	inventory position at the central warehouse at time t
$IP_i(t)$	=	inventory position at retailer i at time t
$IL_0(t)$	=	inventory level at the central warehouse at time t
$IL_i(t)$	=	inventory level at retailer i at time t
$O_i(t)$	=	inventory on hold assigned to retailer i at the central warehouse at time t
S_i	=	expected number of extra deliveries per time unit due to partial deliveries to retailer i
x^+	=	max(x,0)
x^-	=	max(−x,0)

It is assumed that R_0 is an integer multiple of q (and that the central warehouse inventory level is a multiple of q when the system is initiated). As a direct consequence, IL_0 and IP_0 will always be multiples of q. The retailers accept both partial and complete deliveries, but do require information, at the time of order placement, on exactly when they will be receiving the units ordered. Under the MSD policy, this means that the central warehouse is not allowed to revise its decision (e.g., send units on hold earlier than first decided) based on events after the time of retailer order placement. Another consequence of this assumption is that we only consider central warehouse reorder points that satisfy $R_0 \geq 0$. This eliminates situations where there are backorders at the central warehouse, but no order has yet been placed to the outside supplier. As mentioned above, we also assume that partial delivery can lead to a maximum of one extra delivery. This means that we require that $Q_0 \geq \max(Q_1,\ldots,Q_N)$. In most practical situations one would expect the central warehouse order quantity to be larger than the retailer order quantities. Hence, this assumption is not very restrictive. Using our methodology, it is quite easy to expand the parameter range to $R_0 \geq -q$ and $Q_0 \geq \max(Q_1,\ldots,Q_N) - q$, but for ease of exposition we exclude these special cases.

We assume that all order quantities are given (e.g., determined by a deterministic model), and we focus on determining the integer reorder points $\mathbf{R} = (R_0,\ldots,R_N)$ that minimize expected total system costs for each of the three different delivery policies. The expected total costs are given by

$$TC(\mathbf{R}) = h_0(E[IL_0^+] + \sum_{i=1}^{N} E[O_i]) + \sum_{i=1}^{N} \theta_i S_i + \sum_{i=1}^{N}(h_i E[IL_i^+] + b_i E[IL_i^-]). \tag{1}$$

3. Analysis

We begin the analysis with deriving the cost minimizing MSD decision rule. Given this policy, we then provide a method for determining the expected total costs for a given set of reorder points. This is followed by the analysis of the PD policy and the CD policy, respectively. Finally, we provide methods for obtaining the optimal reorder points under each of the three different delivery policies considered.

3.1 The MSD decision rule

Assume that retailer i places an order of Q_i units to the central warehouse at time τ_z. The MSD decision is applied at time τ_z if part of, but not the entire, order will be available for dispatch at some point in time. Because $Q_0 \geq Q_i$ ($\forall\, i \in \mathbf{N}$), the central warehouse never places two orders to the outside supplier at the same time. Clearly, if the central warehouse inventory level is such that $q \leq IL_0(\tau_z) \leq Q_i - q$ just before the retailer order of Q_i is placed, part of the order can be dispatched immediately, and part of the order can be dispatched when the next batch of Q_0 units arrives from the outside supplier. Furthermore, if $q - Q_0 \leq IL_0(\tau_z) \leq Q_i - q - Q_0$, part of the order can be dispatched when the next batch of Q_0 units arrives, and part of the order can be dispatched when the next Q_0 after that arrives. Following this logic the MSD decision is applied if and only if $q - nQ_0 \leq IL_0(\tau_z) \leq Q_i - q - nQ_0$ (n = 0, 1, 2…) when the order of Q_i occurs. If this is not the case, a complete delivery will be made and we do not need to apply the MSD decision. Recall that the order of Q_i can be split in at most two parts. Therefore, for applying the MSD decision at time τ_z we define

t_1 =   time until the first part of the order can be dispatched from the central warehouse ($t_1 \geq 0$)

t_2 =   time until the entire order can be dispatched from the central warehouse ($t_2 > t_1$)

u   =   number of units available for dispatch to retailer i from the central warehouse at time $\tau_z + t_1$ ($q \leq u \leq Q_i - q$).

Recall that $R_0 \geq 0$. Therefore, the values of t_1, t_2 and u will be known at time τ_z. A partial delivery means that we dispatch u units at time $\tau_z + t_1$, and $Q_i - u$ units at time $\tau_z + t_2$. A complete delivery means that we dispatch all Q_i units at time $\tau_z + t_2$. We refer to the latter as placing u units on hold (between time $\tau_z + t_1$ and $\tau_z + t_2$).

The analysis is facilitated by the following observation:

Observation 1

The MSD decision at time τ_z does not affect any forthcoming MSD decisions. Furthermore, the MSD decision only affects the costs at the central warehouse and retailer i.

Observation 1 follows directly from the fact that the MSD decision cannot be revised after time τ_z, and the fact that central warehouse uses FCFS allocation. The FCFS allocation principle implies that retailer orders will be satisfied in the sequence that they arrive at the central warehouse. Therefore, any units on hold for retailer i do not affect other incoming orders at the central warehouse. Observation 1 means that we can derive the decision rule by considering the expected difference in holding - and backorder costs of placing the u units on hold until time $\tau_z + t_2$, compared to dispatching them at time $\tau_z + t_1$. This cost difference, denoted by $\Delta C_i(R_i, t_1, t_2, u)$, can then be compared to the handling cost, θ_i. If $\Delta C_i(R_i, t_1, t_2, u) \leq \theta_i$, complete delivery is chosen. Otherwise, partial delivery is chosen.

We obtain $\Delta C_i(R_i, t_1, t_2, u)$ by analyzing each affected cost component separately. Obviously, placing u units on hold implies an additional holding cost of $h_0(t_2 - t_1)u$ at the central warehouse. To calculate the cost effect at retailer i, we arbitrarily number the u units considered and introduce our second observation:

Observation 2

Given that $R_i + n > 0$ at time τ_z, the nth unit in the u units considered will satisfy the $(R_i + n)^{th}$ customer demand after τ_z at retailer i. If $R_i + n \leq 0$, the nth unit will satisfy the $(R_i + n + 1)^{th}$ most recent backorder before time τ_z.

Observation 2 holds because the retailer's inventory position just reached R_i at time τ_z. Furthermore, the FCFS allocation principle means that the nth unit will satisfy the same customer demand, regardless if it placed on hold at the central warehouse or not. Because customer demand occurs according to a Poisson process, the time until the nth unit is demanded is Erlang distributed. Let

$\Gamma_i(k)$ = stochastic time for k customer demand arrivals at retailer i, $\Gamma_i(k) \in \text{Erlang}(k, \lambda_i)$ with probability density function $f_{\Gamma_i(k)}(x) = \lambda^k x^{k-1} e^{-\lambda x} / (k-1)! \quad x \geq 0$.

$g_i(R_i, \delta, n)$ = expected holding - and backorder cost at retailer i for the nth unit placed on hold, arriving at retailer i in δ time units

It follows that

8

$$\Delta C_i(R_i, t_1, t_2, u) = h_0(t_2 - t_1)u + \sum_{n=1}^{u} g_i(R_i, L_i + t_2, n) - g_i(R_i, L_i + t_1, n). \tag{2}$$

The nth unit will either incur a holding cost or a backorder cost, depending on if it arrives before or after its corresponding demand at the retailer. Hence, for $R_i + n > 0$ we have

$$g_i(R_i, \delta, n) = h_i E\left[\left(\Gamma_i(R_i + n) - \delta\right)^+\right] + b_i E\left[\left(\Gamma_i(R_i + n) - \delta\right)^-\right]$$

$$= (h_i + b_i)E\left[\left(\Gamma_i(R_i + n) - \delta\right)^+\right] + b_i E\left[\left(\delta - \Gamma_i(R_i + n)\right)\right] \tag{3}$$

$$= \frac{(h_i + b_i)e^{-\lambda_i\delta}}{\lambda_i} \sum_{j=0}^{R_i+n} (R_i + n - j)\frac{(\lambda_i\delta)^j}{j!} + \frac{b_i}{\lambda_i}(\lambda_i\delta - R_i - n),$$

and for $R_i + n \leq 0$

$$g_i(R_i, \delta, n) = b_i\delta. \tag{4}$$

The results from our analysis are summarized in Proposition 1.

Proposition 1 *The MSD decision rule*

The central warehouse should choose complete delivery of the batch of Q_i if

$$\Delta C_i(R_i, t_1, t_2, u) \leq \theta_i. \tag{5}$$

Otherwise, partial delivery should be chosen.

Note that if (5) is satisfied with equality, we are indifferent to either partial or complete delivery. Two corollaries follow from Proposition 1.

Corollary 1 *Performance guarantee*

For any set of reorder points, **R**, the MSD policy will always provide an expected total cost which is lower or equal to the PD - and the CD policy.

Corollary 1 follows directly because a cost minimizing decision is made each time an order is placed. It means that the MSD decision rule provides a cost performance guarantee both compared to the PD - and the CD policy, regardless of how the reorder points are chosen. Corollary 2 makes it possible to identify systems where it is always reasonable to use the CD policy.

Corollary 2

If $\theta_i > L_0(Q_i - 1)(h_0 + b_i)$, the MSD policy will always dispatch complete deliveries to retailer i.

Proof

It is sufficient to show that $L_0(Q_i - 1)(h_0 + b_i)$ is an upper bound for $\Delta C_i(t_1,t_2,u)$. This is done by maximizing each cost component separately. The maximum expected central warehouse holding cost difference incurred by placing u units on hold occurs when $u = Q_i - 1$ and $t_2 - t_1 = L_0$. Hence, $\max\{h_0(t_2-t_1)u\} = h_0L_0(Q_i-1)$. The maximum expected retailer holding cost difference is equal to zero (retailer holding costs will never increase by placing units on hold). The maximum expected retailer backorder cost difference is $b_iL_0(Q_i-1)$, which again occurs when $u = Q_i - 1$ and $t_2 - t_1 = L_0$. Thus, $h_0L_0(Q_i-1) + 0 + b_iL_0(Q_i-1) = L_0(Q_i - 1)(h_0 + b_i)$ is an upper bound for $\Delta C_i(t_1,t_2,u)$. ∎

A direct result of Corollary 2 is that the MSD policy will be identical to the CD policy in systems where $\theta_i > L_0(Q_i - 1)(h_0 + b_i)$ for all retailers.

3.2 The MSD policy: Distribution of retailer inventory levels

Recall that in order to obtain the total expected cost for a given set of reorder points **R**, we need to determine $E[IL_0^+]$, $E[O_i]$, S_i, $E[IL_i^+]$ and $E[IL_i^-]$ ($i = 1,\ldots,N$). In this section we derive the probability distributions of the retailer inventory levels, $P(IL_i = m)$. Given $P(IL_i = m)$, it is easy to obtain the expected amount of stock on hand and backorders at retailer i as

$$E[IL_i^+] = \sum_{m=1}^{R_i+Q_i} mP(IL_i = m)$$

$$E[IL_i^-] = \sum_{m=1}^{\infty} mP(IL_i = -m).$$

(6)

Methods for obtaining $E[IL_0^+]$, $E[O_i]$ and S_i are provided in Section 3.3.

It is well known that the retailer inventory positions in steady state are uniformly distributed on the integers $[R_i+1, R_i+2,\ldots, R_i+Q_i]$, and the central warehouse inventory position is uniform on $[R_0+q, R_0+2q,\ldots, R_0+Q_0]$ (e.g. Axsäter 1998). Furthermore, these $N + 1$ distributions are independent. Note that this is true also in our context because the inventory positions are not dependent on the delivery policy. This is because stock on hold is not included in the central warehouse inventory position, and placing units on hold does not change the inventory position of retailer i. Let

$B_i(t)$	=	number of backorders at the central warehouse belonging to retailer i at time t
$D_i(\tau_1,\tau_2)$	=	customer demand at retailer i in the time interval $[\tau_1,\tau_2]$ (Poisson distributed)
$D_0(\tau_1,\tau_2)$	=	demand from all retailers to the central warehouse in the time interval $[\tau_1,\tau_2]$
$D_{0,i}(\tau_1,\tau_2,a,d)$	=	demand from retailer i to the central warehouse in the time interval $[\tau_1,\tau_2]$, given that $IP_i(\tau_1) = a$ and $D_i(\tau_1,\tau_2) = d$

$$\operatorname*{mod}_{R,Q}(z) \quad = \quad z + nQ, \text{ where n is the integer such that } R + 1 \le z + nQ \le R + Q.$$

$D_{0,i}(\tau_1,\tau_2,a,d)$ is easily obtained as nQ_i, where n is the integer such that $R + 1 \le a - d + nQ \le R + Q$.

A unit ordered by retailer i will arrive in stock after L_i time units, unless it has been backordered or placed on hold at the central warehouse. Thus, it must hold for an arbitrary time t that

$$IL_i(t + L_0 + L_i) = IP_i(t + L_0) - B_i(t + L_0) - O_i(t + L_0) - D_i(t + L_0, t + L_0 + L_i). \tag{7}$$

For notational convenience we assume that $t = 0$ and hereafter refer to this arbitrary time as time zero. That is,

$$IL_i(L_0 + L_i) = IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i). \tag{8}$$

$D(L_0, L_0 + L_i)$ is independent of $IP_i(L_0)$, $B_i(L_0)$ and $O_i(L_0)$. However, there clearly are dependencies between $IP_i(L_0)$, $B_i(L_0)$ and $O_i(L_0)$. The remainder of this section focuses on determining the distribution of $IL_i(L_0+L_i)$.

To facilitate the analysis we introduce the nominal inventory position, $\Psi_0(\tau)$, defined for $0 \le \tau \le L_0$. This variable is a modified version of the nominal inventory position introduced in Stenius et al. (2013). The nominal inventory position is defined as the central warehouse inventory position at time zero, minus all retailer demand to the central warehouse in the time interval $[0,\tau]$. That is,

$$\Psi_0(\tau) = IP_0(0) - D_0(0, \tau). \tag{9}$$

$\Psi_0(\tau)$ is a stepwise decreasing stochastic variable containing information about how much demand the central warehouse can satisfy before time L_0. Properties of the nominal inventory position that will prove useful in proceeding sections are stated in Lemma 1.

Lemma 1 *Properties of the nominal inventory position*

1) $\Psi_0(0) = IP_0(0)$

2) $\Psi_0(L_0) = IL_0(L_0)$

   For a given τ ($0 \le \tau \le L_0$):

3) If $\Psi_0(\tau) \ge 0$, no retailer orders occurring before time τ will be backordered or on hold at the central warehouse at time L_0.

4) If $\Psi_0(\tau) \le 0$, all retailer orders that occur in the time interval $(\tau, L_0]$ will be backordered at the central warehouse at time L_0.

5) If $\Psi_0(\tau) = u$ ($q \le u \le Q_i - q$) and a retailer order for Q_i units occurs at time τ, $Q_i - u$ of these units will be backordered at time L_0. The remaining u units will be:

   a) dispatched to the retailer at time L_0 if partial delivery is chosen for the order

   b) on hold at the central warehouse at time L_0 if complete delivery is chosen for the order.

Proof

1) Follows from the definition in (9).

2) Also follows from (9) because $IL_0(L_0) = IP_0(0) - D_0(0,L_0) = \Psi_0(L_0)$.

3) At time L_0, the FCFS assumption means that the central warehouse will have been able to satisfy demand for the first $IP_0(0)$ units in the time interval $[0,L_0]$. Because $\Psi_0(\tau) \geq 0$ means that $D_0(0,\tau) \leq IP_0(0)$, the statement must hold.

4) Must hold because $\Psi_0(\tau) \leq 0$ means that $IP_0(0)$ units have already been satisfied at time τ.

5) When $\Psi_0(\tau) = u$, $IP_0(0) - u$ demands will have occurred just before the order of Q_i occurs at time τ. Thus, u units can be satisfied at time L_0 and $Q_i - u$ units will be backordered. Whether the u units are dispatched or placed on hold at time L_0 follows directly from the MSD decision. ∎

To further facilitate the analysis it is also appropriate to define the system state vector at time τ,

$$\mathbf{V}(\tau) = \left[\Psi_0(\tau), IP_1(\tau),..., IP_N(\tau)\right]. \tag{10}$$

The analysis is based on calculating the probabilities for state transitions of the vector $\mathbf{V}(\tau)$. That is, given a state $\mathbf{V}(\tau_1) = \mathbf{v} = [v_0, v_1,…, v_N]$ at time τ_1, we can calculate the probability of state $\mathbf{V}(\tau_2) = \mathbf{w} = [w_0, w_1,…, w_N]$ at time τ_2 $(\tau_2 \geq \tau_1)$. This leads us to Lemma 2.

Lemma 2

The conditional distribution of $\mathbf{V}(\tau_2)|\mathbf{V}(\tau_1)$ is obtained as

$$P\left(\mathbf{V}(\tau_2) = \mathbf{w} \mid \mathbf{V}(\tau_1) = \mathbf{v}\right) = \sum_{d_1=0}^{\infty} \sum_{d_2=0}^{\infty} ... \sum_{d_N=0}^{\infty} \chi \prod_{i=1}^{N} P\left(D_i(\tau_1,\tau_2) = d_i\right),$$

where $\tag{11}$

$$\chi = \begin{cases} 1 \; ; \; v_0 - \sum_{i=1}^{N} D_{0,i}(\tau_1,\tau_2,v_i,d_i) = w_0, \; \underset{R_1,Q_1}{\mathrm{mod}}(v_1 - d_1) = w_1 ,..., \underset{R_N,Q_N}{\mathrm{mod}}(v_N - d_N) = w_N \\ 0 \; ; \; \text{otherwise.} \end{cases}$$

Proof

Because each retailer faces an independent Poisson process, the probability of a demand realization $d_1,d_2,…,d_N$ in the time interval $[\tau_1,\tau_2]$ is $\prod_{i=1}^{N} P\left(D_i(\tau_1,\tau_2) = d_i\right)$. Furthermore, all different demand realizations are mutually exclusive events. The initial inventory position, v_i, and demand realization, d_i, at retailer i, will lead to a demand for $D_{0,i}(\tau_1,\tau_2,v_i,d_i)$ units at the central warehouse (lowering the nominal inventory position by this amount). The result then follows from the fact that, because of the (R_i,Q_i) policy used at retailer i, the inventory position will be $\underset{R_i,Q_i}{\mathrm{mod}}(v_i - d_i)$ at time τ_2. ∎

Note that the probability distribution in Lemma 2 is written in a general form. In practical calculations, we do not need to consider all d_i because the indicator function χ will be zero for many demand realizations. Furthermore, it is not necessary to consider infinite sums because χ will always be zero above a certain value of d_i.

To determine the distribution of $IL_i(L_0+L_i)$, we condition on an initial state vector at time zero, $\mathbf{V}(0) = \mathbf{a} = [a_0, a_1, \ldots, a_N]$. Note that at time zero $a_0 = IP_0(0)$ and the elements in $\mathbf{V}(0)$ are therefore independent and uniformly distributed. Thus, for retailer i we have that

$$P\big(IL_i(L_0 + L_i) = m\big) = \frac{q}{\prod_{j=0}^{N} Q_j} \sum_{\mathbf{a} \in \mathbf{A}} P\big(IL_i(L_0 + L_i) = m \mid V(0) = \mathbf{a}\big), \tag{12}$$

where the state space $\mathbf{A}$ contains all possible state vectors at time zero. That is,

$$\begin{aligned} \mathbf{A} = \{\mathbf{a} \in \mathbb{Z}^{N+1} \mid a_0 &= R_0 + q, R_0 + 2q, \ldots, R_0 + Q_0; \\ a_n &= R_n + 1, R_n + 2, \ldots, R_N + Q_N, \forall n \in \mathbf{N}\}. \end{aligned} \tag{13}$$

In light of Lemma 1 we consider three mutually exclusive and (conditioned on $\mathbf{V}(0) = \mathbf{a}$) collectively exhaustive events:

I. Conditioned on $\mathbf{V}(0) = \mathbf{a}$, the nominal inventory position is positive at time L_0.

II. Conditioned on $\mathbf{V}(0) = \mathbf{a}$, the nominal inventory position is non-positive at time L_0, and was brought from a positive to a non-positive value by an order from retailer $j \neq i$ in the time interval $(0, L_0]$.

III. Conditioned on $\mathbf{V}(0) = \mathbf{a}$, the nominal inventory position is non-positive at time L_0, and was brought from a positive to a non-positive value by an order from retailer i in the time interval $(0, L_0]$.

Focusing on retailer i, it follows from Lemma 1 that it is only in Event III that we need to take the MSD decision into consideration. The conditional probability in (12) can now be expressed as

$$\begin{aligned} &P\big(IL_i(L_0 + L_i) = m \mid V(0) = \mathbf{a}\big) \\ &= P\big(IL_i(L_0 + L_i) = m, I\big) + P\big(IL_i(L_0 + L_i) = m, II\big) + P\big(IL_i(L_0 + L_i) = m, III\big). \end{aligned} \tag{14}$$

In the following subsections we show how to determine each term in (14) separately.

3.2.1 Probability of inventory level m in Event I
Given $\mathbf{V}(0) = \mathbf{a}$, the nominal inventory position is positive at time L_0. This means that there will be no units backordered or on hold at the central warehouse at time L_0 (see Lemma 1). Figure 1 depicts a

possible sample path of the nominal inventory position in Event I. The system state at time L_0 is $\mathbf{V}(L_0) = \boldsymbol{\omega} \in \boldsymbol{\Omega}$, where

$$\boldsymbol{\Omega} = \{\boldsymbol{\omega} \in \mathbb{Z}^{N+1} \mid \omega_0 = q, 2q, \ldots, R_0 + Q_0; \qquad (15)$$
$$\omega_n = R_n + 1, R_n + 2, \ldots, R_n + Q_n, \forall n \in \mathbf{N}\}.$$

Lemma 3 provides the key to obtaining the distribution of the inventory level.
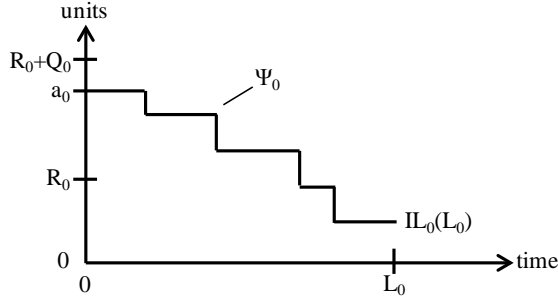
Figure 1. Possible sample path of the nominal inventory position in Event I.

Lemma 3

For $\boldsymbol{\omega} \in \boldsymbol{\Omega}$ we have

$$P\big(IL_i(L_0 + L_i) = m, I \mid \mathbf{V}(L_0) = \boldsymbol{\omega}\big) = P\big(D_i(L_0, L_0 + L_i) = \omega_i - m\big). \qquad (16)$$

Proof

It is given that $IP_i(L_0) = \omega_i$. Furthermore, because $\omega_0 = \Psi_0(L_0) > 0$, Lemma 1 gives that $B_i(L_0) = O_i(L_0) = 0$. Thus, in Event I and conditioned on $\mathbf{V}(L_0) = \boldsymbol{\omega}$ we have

$$IL_i(L_0 + L_i) = IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i)$$
$$= \omega_i - D_i(L_0, L_0 + L_i) \qquad (17)$$

Hence, $IL_i(L_0+L_i) = m$ if and only if $D_i(L_0,L_0+L_i) = \omega_i - m$. ∎

Given Lemma 3, obtaining the distribution of $IL_i(L_0+L_i)$ is simply a matter of considering all state transitions from the vector $\mathbf{a}$ to all vectors in $\boldsymbol{\Omega}$. That is, using the law of total probability we have

$$P\big(IL_i(L_0 + L_i) = m, I\big) = \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} P\big(D_i(L_0, L_0 + L_i) = \omega_i - m\big) P\big(\mathbf{V}(L_0) = \boldsymbol{\omega} \mid \mathbf{V}(0) = \mathbf{a}\big), \qquad (18)$$

where the state transition probabilities are provided in Lemma 2. Note that, because not all state transitions considered in (18) are possible, $P(\mathbf{V}(L_0) = \boldsymbol{\omega} \mid \mathbf{V}(0) = a)$ will be equal to zero in many cases. For instance, because the nominal inventory position is decreasing, ω_0 can never be larger than a_0. However, for ease of exposition we maintain this general notation throughout the proceeding sections.

3.2.2 Probability of inventory level m in Event II

Given $\mathbf{V}(0) = \mathbf{a}$, the nominal inventory position is non-positive at time L_0, and it was an order placed by retailer $j \neq i$ that brought the nominal inventory position to a non-positive value. All orders placed after this particular order by retailer $j \neq i$ will be backordered, and there will be no units on hold to retailer i at time L_0 (see Lemma 1). To determine the distribution of the inventory level, we study the time at which the order by retailer $j \neq i$ occurs. Let

τ_w =   time when a demand occurs at retailer $j \neq i$, triggering an order that brings the nominal

            inventory position from a positive to a non-positive value, $0 < \tau_w \leq L_0$.

We use the notation $\tau_.^{(-)}$ and $\tau_.^{(+)}$ to denote the time just before and just after time $\tau_.$, respectively. That is, $\tau_w^{(-)} < \tau_w < \tau_w^{(+)}$, where $\tau_w^{(+)} = \tau_w^{(-)} + d\tau_w$, $d\tau_w \to 0$. Figure 2 depicts a possible sample path of the nominal inventory position in Event II.
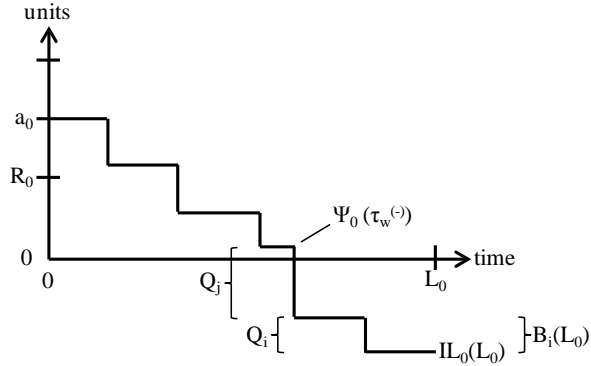
Figure 2. Possible sample path of the nominal inventory position in Event II.

In the example in Figure 2, retailer i places one order after the nominal inventory position has become negative, resulting in Q_i backorders at time L_0.

At time $\tau_w^{(-)}$ the nominal inventory position is positive and less than or equal to Q_j, the inventory position of retailer j is $R_j + 1$, and the inventory positions at all other retailers can be any possible values. Hence, the system state at this time is $\mathbf{V}(\tau_w^{(-)}) = \mathbf{w} \in \mathbf{W}_j^{(-)}$, where

$$\mathbf{W}_j^{(-)} = \{\mathbf{w} \in \mathbb{Z}^{N+1} \mid w_0 = q, 2q, \ldots, Q_j; \; w_j = R_j + 1;$$
$$w_n = R_n + 1, R_n + 2, \ldots, R_n + Q_n, \forall n \in \mathbf{N} \setminus j\}. \tag{19}$$

Conditioning on the state vector $\mathbf{w}$ and on that a demand at retailer j occurs in the time interval $(\tau_w^{(-)}, \tau_w^{(+)})$. Lemma 4 provides the necessary result for analyzing the inventory level of retailer i.

Lemma 4

For $j \in \mathbf{N}\backslash i$ and $\mathbf{w} \in \mathbf{W}_j^{(-)}$ we have

$$P\left(IL_i(L_0 + L_i) = m, II \mid \mathbf{V}(\tau_w^{(-)}) = \mathbf{w}, D_j(\tau_w^{(-)}, \tau_w^{(+)}) = 1\right) = P\left(D_i(\tau_w^{(+)}, L_0 + L_i) = w_i - m\right). \quad (20)$$

Proof

A demand occurred at retailer j in the time interval $(\tau_w^{(-)}, \tau_w^{(+)})$. Because the probability of two demand occurrences in this interval of length $d\tau_w \to 0$ is zero, it follows that $IP_i(\tau_w^{(-)}) = IP_i(\tau_w^{(+)}) = w_i$. Hence, the inventory position at time L_0 must be $IP_i(L_0) = \underset{R,Q}{\mathrm{mod}}\left(w_i - D_i(\tau_w^{(+)}, L_0)\right)$ $= w_i - D_i(\tau_w^{(+)}, L_0) + D_{0,i}(\tau_w^{(+)}, L_0, w_i, D_i(\tau_w^{(+)}, L_0))$. Because $\Psi_0(\tau_w^{(+)}) \leq 0$ due to an order by retailer j, it follows from Lemma 1 that $B_i(L_0) = D_{0,i}(\tau_w^{(+)}, L_0, w_i, D_i(\tau_w^{(+)}, L_0))$ and $O_i(L_0) = 0$. Therefore, in this case we have

$$\begin{aligned}
IL_i(L_0 + L_i) &= IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i) \\
&= w_i - D_i(\tau_w^{(+)}, L_0) + D_{0,i}(\tau_w^{(+)}, L_0, w_i, D_i(\tau_w^{(+)}, L_0)) \\
&\qquad - D_{0,i}(\tau_w^{(+)}, L_0, w_i, D_i(\tau_w^{(+)}, L_0)) - D_i(L_0, L_0 + L_i) \\
&= w_i - D_i(\tau_w^{(+)}, L_0 + L_i)
\end{aligned} \quad (21)$$

Hence, $IL_i(L_0 + L_i) = m$ if and only if $D_i(\tau_w^{(+)}, L_0 + L_i) = w_i - m$. ∎

Lemma 4 shows that conditioned on the event that a state transition from $\mathbf{a}$ to $\mathbf{w}$ occurs, and that the nominal inventory position is brought to a non-positive value by retailer j in the time interval $(\tau_w^{(-)}, \tau_w^{(+)})$, the distribution of the inventory level is a Poisson probability. It remains to determine the probability for such an event to occur, in other words, to determine the joint distribution of $\mathbf{V}(\tau_w^{(-)}) = \mathbf{w}$ and $D_j(\tau_w^{(-)}, \tau_w^{(+)})) = 1$, conditioned on $\mathbf{V}(0) = \mathbf{a}$. The probability of a system state transition from $\mathbf{V}(0) = \mathbf{a}$ to $\mathbf{V}(\tau_w^{(-)}) = \mathbf{w}$, i.e. $P(\mathbf{V}(\tau_w^{(-)}) = \mathbf{w} \mid \mathbf{V}(0) = \mathbf{a})$, is given by Lemma 2. Because of the Poisson demand process, demand in the time interval $(\tau_w^{(-)}, \tau_w^{(+)})$ at retailer j is independent of $\mathbf{V}(\tau_w^{(-)})$, and the probability of exactly one occurrence is $\lambda_j d\tau_w$. The probability of more than one occurrence is zero. Through Lemma 4 we can now determine the distribution of $IL_i(L_0 + L_i)$ using the law of total probability. Summation over all possible state vectors in $\mathbf{W}_j^{(-)}$, over all retailers except retailer i, and over all time intervals of length $d\tau_w$ (i.e., integration) thus yields

$$P\left(IL_i(L_0 + L_i) = m, II\right) =$$
$$\int_0^{L_0} \sum_{j \in \mathbf{N}\backslash i} \sum_{\mathbf{w} \in \mathbf{W}_j^{(-)}} P\left(D_i(\tau_w, L_0 + L_i) = w_i - m\right) P\left(\mathbf{V}(\tau_w) = \mathbf{w} \mid \mathbf{V}(0) = \mathbf{a}\right) \lambda_j d\tau_w. \quad (22)$$

3.2.3 Probability of inventory level m in Event III

Assume that retailer i places an order to the central warehouse at time τ_z ($0 < \tau_z \leq L_0$), and that this brings the nominal inventory position from a positive value u ($q \leq u \leq Q_i$) to a non-positive value $u - Q_i$. If u is strictly less than Q_i, the MSD decision will be applied. This means that the u units will either have been dispatched (partially delivered) or they will be on hold at the central warehouse at time L_0 (see Lemma 1). Because of the FCFS allocation policy, the first u units will be satisfied by the last order placed by the central warehouse before time zero (referred to as Order 1). The remaining $Q_i - u$ units will be satisfied by the first warehouse order placed after time zero (referred to as Order 2). To make the MSD decision we therefore need to keep track of the times when these two warehouse orders were placed (and thus the times when they will be available at the central warehouse). Similar to Event II we define:

τ_x = time when a demand occurs at retailer k = 1,…,N, triggering an order that, in turn, triggers Order 1 at the central warehouse $\tau_x < 0$

τ_y = time when a demand occurs at retailer j = 1,…,N, triggering an order that, in turn, triggers Order 2 at the central warehouse, $0 < \tau_y \leq L_0$

τ_z = time when a demand occurs at retailer i triggering an order that brings the nominal inventory position from a positive to a non-positive value, $\tau_y \leq \tau_z \leq L_0$.

We will treat the event where $\tau_y = \tau_z$ separately (the event where it is the same order from retailer i that both triggers Order 2 and brings the nominal inventory position to a non-positive value). Thus, first assume that $\tau_y < \tau_z$. We refer to this as Event IIIa and the former as Event IIIb, where

$$P\big(IL_i(L_0 + L_i) = m, III\big) = P\big(IL_i(L_0 + L_i) = m, IIIa\big) + P\big(IL_i(L_0 + L_i) = m, IIIb\big). \qquad (23)$$

In order to characterize the system state at time τ_x, we need to extend the definition of the nominal inventory position to include times before time zero. Hence, the nominal inventory position at time $\tau \leq 0$ is defined as the central warehouse inventory position at time zero, plus all retailer demand to the central warehouse in the time interval $[\tau,0]$. That is,

$$\Psi_0(\tau) = IP_0(0) + D_0(\tau,0). \qquad (24)$$

Figure 3 depicts a possible sample path of the nominal inventory position in Event IIIa.
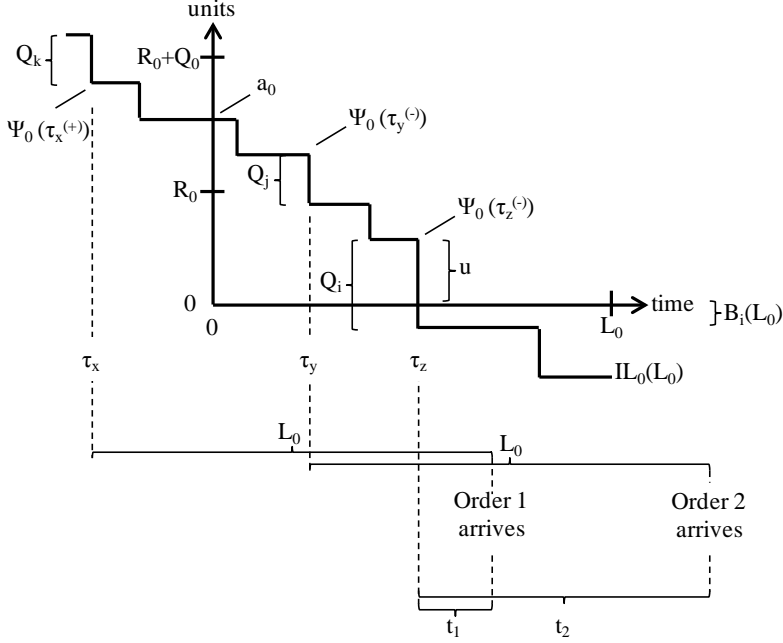
Figure 3. Possible sample path of the nominal inventory position in Event IIIa.

As illustrated in Figure 3, Order 1 is triggered at time τ_x. Just after this event, at time $\tau_x^{(+)}$, the system is in a state such that a demand at retailer k (k = 1,2,…,N) just triggered orders both at that retailer and at the central warehouse. Note that with the extended definition of the nominal inventory position, Order 1 is triggered when Ψ_0 moves from a value above $R_0 + Q_0$ to a value equal to, or below, $R_0 + Q_0$. Also, the inventory position at the retailer that just placed the order, retailer k, must be $R_k + Q_k$. This means that $\mathbf{V}(\tau_x^{(+)}) = \mathbf{x} \in \mathbf{X}_k^{(+)}$, where

$$\mathbf{X}_k^{(+)} = \{\mathbf{x} \in \mathbb{Z}^{N+1} \mid x_0 = R_0 + Q_0 - Q_k + q, R_0 + Q_0 - Q_k + 2q,\ldots, R_0 + Q_0;\ x_k = R_k + Q_k; \tag{25}$$
$$x_n = R_n + 1, R_n + 2,\ldots, R_n + Q_n, \forall n \in \mathbf{N} \setminus k\}.$$

At time $\tau_y^{(-)}$, a demand at retailer j (j = 1,2,…,N) is about to trigger Order 2, but the nominal inventory position will remain positive (because we are considering Event IIIa). Hence, $\mathbf{V}(\tau_y^{(-)}) = \mathbf{y} \in \mathbf{Y}_j^{(-)}$,

$$\mathbf{Y}_j^{(-)} = \{\mathbf{y} \in \mathbb{Z}^{N+1} \mid y_0 = \max(R_0, Q_j) + q, \max(R_0, Q_j) + 2q,\ldots, R_0 + Q_j;\ y_j = R_j + 1; \tag{26}$$
$$y_n = R_n + 1, R_n + 2,\ldots, R_n + Q_n, \forall n \in \mathbf{N} \setminus j\}.$$

The moment after $\tau_y^{(-)}$, at time $\tau_y^{(+)}$, the nominal inventory position will have decreased by Q_j, and the inventory position of retailer j will have increased by $Q_j - 1$. That is, $\mathbf{V}(\tau_y^{(+)}) = f_j(\mathbf{y})$, where

$$f_j(\mathbf{y}) = (y_0 - Q_j, y_1,\ldots, y_j - 1 + Q_j,\ldots, y_N),\quad \mathbf{y} \in \mathbf{Y}_j^{(-)}. \tag{27}$$

At time $\tau_z^{(-)}$ retailer i is just about to place an order that brings the nominal inventory position from a positive to a non-positive value. We have $\mathbf{V}(\tau_z^{(-)}) = \mathbf{z} \in \mathbf{Z}_i^{(-)}$,

18

$$\mathbf{Z}_i^{(-)} = \{\mathbf{z} \in \mathbb{Z}^{N+1} \mid z_0 = q, 2q, \ldots, Q_i; \ z_i = R_i + 1;$$
$$z_n = R_n + 1, R_n + 2, \ldots, R_n + Q_n, \forall n \in \mathbf{N} \setminus i\}. \tag{28}$$

Figure 3 shows the relation between τ_x, τ_y, τ_z and the input parameters t_1 and t_2 that are used for the MSD decision. We have that $t_1 = (\tau_x + L_0 - \tau_z)^+$ and $t_2 = \tau_y + L_0 - \tau_z$. The number of units available for partial delivery, u, is given by the value of the nominal inventory position at time $\tau_z^{(-)}$. That is, $u = z_0$.

Similar to Lemma 4, we condition on the event that Order 1 is triggered in $(\tau_x^{(-)}, \tau_x^{(+)})$, Order 2 is triggered in $(\tau_y^{(-)}, \tau_y^{(+)})$, and the nominal inventory position is brought to a non-positive value in $(\tau_z^{(-)}, \tau_z^{(+)})$. We have Lemma 5:

Lemma 5

For $k \in \mathbf{N}$, $j \in \mathbf{N}$, $\mathbf{x} \in \mathbf{X}_k^{(+)}$, $\mathbf{y} \in \mathbf{Y}_k^{(-)}$ and $\mathbf{z} \in \mathbf{Z}_j^{(-)}$ we have

$$P\Big(IL_i(L_0 + L_i) = m, IIIa \mid D_k(\tau_x^{(-)}, \tau_x^{(+)}) = 1, \mathbf{V}(\tau_x^{(+)}) = \mathbf{x},$$
$$\mathbf{V}(\tau_y^{(-)}) = \mathbf{y}, D_j(\tau_y^{(-)}, \tau_y^{(+)}) = 1, \mathbf{V}(\tau_z^{(-)}) = \mathbf{z}, D_i(\tau_z^{(-)}, \tau_z^{(+)}) = 1\Big) = \tag{29}$$
$$\begin{cases} P\Big(D_i(\tau_z^{(+)}, L_0 + L_i) = R_i - m\Big) & ; \ u < Q_i \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) \le \theta_i \\ P\Big(D_i(\tau_z^{(+)}, L_0 + L_i) = R_i + u - m\Big); \ \text{otherwise} \end{cases},$$

where $u = z_0$, $t_1 = (\tau_x + L_0 - \tau_z)^+$ and $t_2 = \tau_y + L_0 - \tau_z$.

Proof

Because an order was just triggered at retailer i at time $\tau_z^{(+)}$ we have $IP_i(\tau_z^{(+)}) = R_i + Q_i$. Therefore, $IP_i(L_0) = R_i + Q_i - D_i(\tau_z^{(+)}, L_0) + D_{0,i}(\tau_z^{(+)}, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0))$. An order occurs just after time $\tau_z^{(-)}$ that brings the nominal inventory position from a positive value, $\Psi(\tau_z^{(-)}) = z_0 = u$, to a non-positive value $\Psi(\tau_z^{(-)}) = u - Q_i$. Lemma 2 thus implies that $B_i(L_0) = Q_i - u + D_{0,i}(\tau_z^{(+)}, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0))$.

The amount of units on hold at time L_0, $O(L_0)$, depends on the MSD decision (unless $u = Q_i$ in which case no decision is made and $O_i(L_0) = 0$). Order 1 was placed at time $\tau_x^{(+)}$ meaning that this order will arrive at the central warehouse at time $\tau_x^{(+)} + L_0$. Correspondingly, Order 2 will arrive at time $\tau_y^{(+)} + L_0$. Hence, when applying the MSD decision rule at time $\tau_z^{(+)}$ the first u units will be available after $t_1 = (\tau_x^{(+)} + L_0 - \tau_z^{(+)})^+$ time units and the remaining $Q_i - u$ units will be available after $t_2 = \tau_y^{(+)} + L_0 - \tau_z^{(+)}$ time units. If $u < Q_i$, the MSD decision is applied and Proposition 1 gives that $O_i(L_0) = u$, if $\Delta C_i(R_i, T_1, T_2, u) \le \theta_i$. Otherwise $O_i(L_0) = 0$. In the former case (complete delivery) we therefore have

$$IL_i(L_0 + L_i) = IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i)$$
$$= R_i + Q_i - D_i(\tau_z^{(+)}, L_0) + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0))$$
$$- \left(Q_i - u + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0)) \right) - u - D_i(L_0, L_0 + L_i) \tag{30}$$
$$= R_i - D_i(\tau_z^{(+)}, L_0 + L_i),$$

and in the latter case (partial delivery or if $u = Q_i$) we have

$$IL_i(L_0 + L_i) = IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i)$$
$$= R_i + Q_i - D_i(\tau_z^{(+)}, L_0) + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0))$$
$$- \left(Q_i - u + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0)) \right) - D_i(L_0, L_0 + L_i) \tag{31}$$
$$= R_i + u - D_i(\tau_z^{(+)}, L_0 + L_i).$$

∎

It remains to determine the probability of the event conditioned on in Lemma 5. Analogously to Event II, all state transitions depend only on the starting state and the customer demand in disjoint time intervals. We can therefore again obtain the joint distribution by multiplying all the probabilities of moving from one state to the other. That is, we determine the probability of the sample path where: (i) Order 1 is triggered in $(\tau_x^{(-)}, \tau_x^{(+)})$ by an order from retailer k, resulting in the state space $\mathbf{V}(\tau_x^{(+)}) = \mathbf{x}$, (ii) the state at time zero is $\mathbf{V}(0) = \mathbf{a}$, (iii) Order 2 is triggered in $(\tau_y^{(-)}, \tau_y^{(+)})$ from the state $\mathbf{V}(\tau_y^{(-)}) = \mathbf{y}$ by an order from retailer j, and (iv) the nominal inventory position is moved to a non-positive value in $(\tau_z^{(-)}, \tau_z^{(+)})$ from state $\mathbf{V}(\tau_z^{(-)}) = \mathbf{z}$ by an order from retailer i. This yields

$$\lambda_k d\tau_x P\left(\mathbf{V}(0) = \mathbf{a} \mid \mathbf{V}(\tau_x^{(+)}) = \mathbf{x} \right) P\left(\mathbf{V}(\tau_y^{(-)}) = \mathbf{y} \mid \mathbf{V}(0) = \mathbf{a} \right) \lambda_j d\tau_y P\left(\mathbf{V}(\tau_z^{(-)}) = \mathbf{z} \mid \mathbf{V}(0) = f_j(\mathbf{y}) \right) \lambda_i d\tau_z. \tag{32}$$

Utilizing Lemma 5 and considering all possible times, retailers and state vectors we have

$$P\left(IL_i(L_0 + L_i) = m, IIIa \right) =$$
$$\int_{-\infty}^{0} \int_{0}^{L_0} \int_{\tau_y}^{L_0} \sum_{k \in \mathbf{N}} \sum_{j \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\mathbf{y} \in \mathbf{Y}_j^{(-)}} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} p_i(m, \tau_x, \tau_y, \tau_z, \mathbf{z}) \, \pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) d\tau_z d\tau_y d\tau_x, \tag{33}$$

where $p_i()$ is given by Lemma 5,

$$p_i(m, \tau_x, \tau_y, \tau_z, \mathbf{z}) =$$
$$\begin{cases} P\left(D_i(\tau_z, L_0 + L_i) = R_i - m \right) & ; \quad u < Q_i \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) \leq \theta_i \\ P\left(D_i(\tau_z, L_0 + L_i) = R_i + u - m \right) & ; \quad \text{otherwise} \end{cases} \tag{34}$$
$$u = z_0, t_1 = (\tau_x + L_0 - \tau_z)^+, t_2 = \tau_y + L_0 - \tau_z,$$

and

$$\pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) =$$
$$\lambda_k P\left(\mathbf{V}(0) = \mathbf{a} \mid \mathbf{V}(\tau_x) = \mathbf{x} \right) P\left(\mathbf{V}(\tau_y) = \mathbf{y} \mid \mathbf{V}(0) = \mathbf{a} \right) \lambda_j P\left(\mathbf{V}(\tau_z) = \mathbf{z} \mid \mathbf{V}(0) = f_j(\mathbf{y}) \right) \lambda_i \tag{35}$$

is obtained from (32).

We now turn to Event IIIb, where it is the same order from retailer i that both triggers Order 2 and brings the nominal inventory position to a non-positive value. Note that $R_0 \leq Q_i - q$ for this to be able to occur. Figure 4 depicts a possible sample path of the nominal inventory position in this event.


Figure 4. Possible sample path of the nominal inventory position in Event IIIb.

The analysis is simpler than the preceding case as we do not need to consider two separate times $\tau_y^{(-)}$ and $\tau_z^{(-)}$, or sets $Y_j^{(-)}$ and $Z_j^{(-)}$. The state just before a demand at retailer i occurs, triggering an order to the central warehouse, which simultaneously triggers Order 2 and brings the nominal inventory position to a non-positive value, is $\mathbf{V}(\tau_y^{(-)}) = \tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}$, where

$$\tilde{\mathbf{Y}}_i^{(-)} = \{\tilde{\mathbf{y}} \in \mathbb{Z}^{N+1} \mid \tilde{y}_0 = R_0 + q, R_0 + 2q \ldots, Q_i; \ \tilde{y}_i = R_i + 1;$$
$$\tilde{y}_n = R_n + 1, R_n + 2, \ldots, R_n + Q_n, \forall n \in \mathbf{N} \setminus i\}. \tag{36}$$

Following the same analysis as above we have

$$P\big(IL_i(L_0 + L_i) = m, IIIb\big) =$$
$$\int_{-\infty}^{0} \int_{0}^{L_0} \sum_{k \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}} \tilde{p}_i(m, \tau_x, \tau_y, \tilde{\mathbf{y}}) \tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) d\tau_y d\tau_x, \tag{37}$$

with

$$\tilde{p}_i(m, \tau_x, \tau_y, \tilde{\mathbf{y}}) =$$

$$\begin{cases} P\big(D_i(\tau_y, L_0 + L_i) = R_i - m\big) & ; \quad u < Q_i \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) \le \theta_i \\ P\big(D_i(\tau_y, L_0 + L_i) = R_i + u - m\big) & ; \quad \text{otherwise} \end{cases} \tag{38}$$

$$u = \tilde{y}_0, t_1 = (\tau_x + L_0 - \tau_y)^+, t_2 = L_0,$$

and

$$\tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) =$$

$$\lambda_k P\big(\mathbf{V}(0) = \mathbf{a} \mid \mathbf{V}(\tau_x) = \mathbf{x}\big) P\big(\mathbf{V}(\tau_y) = \tilde{\mathbf{y}} \mid \mathbf{V}(0) = \mathbf{a}\big)\lambda_i. \tag{39}$$

3.3 Determining $E[O_i]$, S_i and $E[IL_0^+]$ under the MSD policy

The evaluation of the expected number of units on hold to retailer i, $E[O_i]$, the expected number of extra deliveries per time unit due to partial deliveries to retailer i, and the expected number of units on hand at the central warehouse, $E[IL_0^+]$, are all based on the analysis in Section 3.2. Note that given the distributions of $O_i(L_0)$ and $IL_0^+(L_0)$, the expected values are obtained as

$$E[O_i] = \sum_{u=q}^{Q_i - q} u P\big(O_i(L_0) = u\big), \tag{40}$$

$$E\Big[IL_0^+\Big] = \sum_{m=q}^{R_0 + Q_0} m P\big(IL_0(L_0) = m\big). \tag{41}$$

3.3.1 Units on hold, $O_i(L_0)$

Based on the analysis in Section 3.2, we know that it is only in Event III that there can be units on hold to retailer i at time L_0. The u units on hold always belong to the same order, with $u = q, 2q, \ldots, Q_i - q$. Defining $\mathbf{A}$ as in (13) and Events IIIa and IIIb (which are conditioned on $\mathbf{V}(0) = \mathbf{a}$) as in Section 3.2.3, gives the unconditioned probability

$$P\big(O_i(L_0) = u\big) = \begin{cases} \dfrac{q}{\prod\limits_{j=0}^{N} Q_j} \sum\limits_{\mathbf{a} \in \mathbf{A}} P\big(O_i(L_0) = u, \text{IIIa}\big) + P\big(O_i(L_0) = u, \text{IIIb}\big); & u = q, 2q, \ldots Q_i - q \\ 0 & ; \quad \text{otherwise.} \end{cases} \tag{42}$$

In order for $P(O_i(L_0) = u) > 0$ in Event IIIa, the nominal inventory position at time $\tau_z^{(-)}$ must be exactly u units. Thus, we have $\mathbf{V}(\tau_z^{(-)}) = \mathbf{z} \in \mathbf{Z_i}^{(-)}(\mathbf{u})$,

$$\mathbf{Z}_i^{(-)}(u) = \{\mathbf{z} \in \mathbb{Z}^{N+1} \mid z_0 = u; \ z_i = R_i + 1;$$
$$z_n = R_n + 1, R_n + 2, \ldots, R_n + Q_n, \forall n \in \mathbf{N} \setminus i\}, \tag{43}$$

for $u = q, 2q, \ldots, Q_i - q$, and $\mathbf{Z}_i^{(-)}(\mathbf{u}) = \varnothing$ otherwise. Following the same logic as in Section 3.2.3 yields

$$P\left(O_i(L_0) = u, IIIa\right) =$$

$$\int_{-\infty}^{0} \int_{0}^{L_0} \int_{\tau_y}^{L_0} \sum_{k \in N} \sum_{j \in N} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\mathbf{y} \in \mathbf{Y}_j^{(-)}} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}(u)} \rho_i(\tau_x, \tau_y, \tau_z, u) \, \pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) d\tau_z d\tau_y d\tau_x, \tag{44}$$

where

$$\rho_i(\tau_x, \tau_y, \tau_z, u) = \begin{cases} 1 & ; \quad \Delta C_i(R_i, t_1, t_2, u) \le \theta_i \\ 0 & ; \quad \text{otherwise,} \end{cases} \tag{45}$$

$t_1 = (\tau_x + L_0 - \tau_z)^+$, $t_2 = \tau_y + L_0 - \tau_z$, and π_{kji} is obtained from (35).

Analogously, for Event IIIb we have $\mathbf{V}(\tau_y^{(-)}) = \tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}(u)$,

$$\tilde{\mathbf{Y}}_i^{(-)}(u) = \{\tilde{\mathbf{y}} \in \mathbb{Z}^{N+1} \mid \tilde{y}_0 = u; \ \tilde{y}_i = R_i + 1; \\ \tilde{y}_n \le R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus i\}, \tag{46}$$

for $u = R_i + q, R_i + 2q, \dots, Q_i - q$ and $\tilde{\mathbf{Y}}_i^{(-)}(u) = \varnothing$ otherwise. This results in

$$P\left(O_i(L_0) = u, IIIb\right) =$$

$$\int_{-\infty}^{0} \int_{0}^{L_0} \sum_{k \in N} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_j^{(-)}(u)} \tilde{\rho}_i(\tau_x, \tau_y, u) \, \tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) d\tau_y d\tau_x, \tag{47}$$

where

$$\tilde{\rho}_i(\tau_x, \tau_y, u) = \begin{cases} 1 & ; \quad \Delta C_i(R_i, t_1, t_2, u) \le \theta_i \\ 0 & ; \quad \text{otherwise,} \end{cases} \tag{48}$$

$t_1 = (\tau_x + L_0 - \tau_z)^+$, $t_2 = L_0$, and $\tilde{\pi}_{ki}$ is obtained from (39).

3.3.2 Rate of partial deliveries, S_i

As stated in the problem formulation, a cost θ_i is incurred each time an order is shipped partially to retailer i. We now turn our attention to this cost and determine the expected rate at which partial delivery decisions to retailer i are made, S_i. We only consider $\theta_i > 0$, because if $\theta_i = 0$ the value of S_i does not affect the total costs. The analysis is based on calculating in steady state: (i) the probability that an order to retailer i is partially delivered at time L_0 (i.e., the probability that the first $u < Q_i$ units have been shipped, and the remaining $Q_i - u$ units have not been shipped, at time L_0), (ii) the time difference between the shipping of the first and second part of the given order, which is referred to as the split time and is denoted by T. Using the previous notation it follows that $T = t_2 - t_1$.

Let f(T) be the probability that there is a partially delivered order at time L_0 with the split time T (more precisely, the split time is in the interval $(T^{(-)}, T^{(+)})$). It follows that the expected rate of partial delivery decisions for orders with split time T is f(T) / T, and hence that

$$S_i = \int_{\varepsilon}^{L_0} \frac{f(T)}{T} dT, \tag{49}$$

where ε $(0 < \varepsilon < L_0)$ is a lower bound on T for when the MSD policy will choose partial delivery. Following the same reasoning as for Corollary 2, it is straightforward to show that one such bound is $\varepsilon = \theta_i \, / \, [(h_0 + b_i)(Q_i - 1)]$. If $\varepsilon \geq L_0$ we have $S_i = 0$.

It remains to determine the distribution of f(T). Once again note that there can only be a partially delivered order at time L_0 in Event IIIa and Event IIIb. Thus, utilizing the results in the previous analysis, considering all states $\mathbf{V}(0) = \mathbf{a}$ and all situations where the MSD policy chooses partial delivery, we have

$$f(T) = \frac{q}{\prod\limits_{j=0}^{N} Q_j} \sum_{\mathbf{a} \in \mathbf{A}}$$

$$\int\limits_{-\infty}^{0} \int\limits_{0}^{L_0} \int\limits_{\tau_y}^{L_0} \sum_{k \in \mathbf{N}} \sum_{j \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\mathbf{y} \in \mathbf{Y}_j^{(-)}} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} \sigma_i(\tau_x, \tau_y, \tau_z, \mathbf{z}, T)\, \pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) d\tau_z d\tau_y d\tau_x \; + \quad\quad (50)$$

$$\int\limits_{-\infty}^{0} \int\limits_{0}^{L_0} \sum_{k \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}} \tilde{\sigma}_i(\tau_x, \tau_y, \tilde{\mathbf{y}}, T)\, \tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) d\tau_y d\tau_x,$$

where

$$\sigma_i(\tau_x, \tau_y, \tau_z, \mathbf{z}, T) = \begin{cases} 1 & ; \quad u < Q_i, \quad t_2 - t_1 = T \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) > \theta_i \\ 0 & ; \quad \text{otherwise,} \end{cases} \quad\quad (51)$$

$u = z_0$, $t_1 = (\tau_x + L_0 - \tau_z)^+$, $t_2 = \tau_y + L_0 - \tau_z$, and

$$\tilde{\sigma}_i(\tau_x, \tau_y, \tilde{\mathbf{y}}) = \begin{cases} 1 & ; \quad u < Q_i \quad t_2 - t_1 = T \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) > \theta_i \\ 0 & ; \quad \text{otherwise,} \end{cases} \quad\quad (52)$$

$u = y_0$, $t_1 = (\tau_x + L_0 - \tau_y)^+$, $t_2 = L_0$. Again, $\pi_{kji}()$ and $\tilde{\pi}_{ki}()$ are obtained from (35) and (39), respectively.

3.3.3 Inventory level at central warehouse, $IL_0(L_0)$

The distribution of the central warehouse inventory level is obtained by analyzing the system at time L_0. In order for $IL_0(L_0) = m$, the system state must be $\mathbf{V}(L_0) = \boldsymbol{\omega} \in \Omega(m)$, where

$$\boldsymbol{\Omega}(m) = \{ \boldsymbol{\omega} \in \mathbb{Z}^{N+1} \mid \omega_0 = m; \\ \omega_n = R_n + 1, R_n + 2, \ldots, R_n + Q_n, \forall n \in \mathbf{N} \}. \quad\quad (53)$$

The probability for this is obtained as

$$P\big(IL_0(L_0) = m\big) = \frac{q}{\prod\limits_{j=0}^{N} Q_j} \sum_{\mathbf{a} \in \mathbf{A}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}(m)} P\big(\mathbf{V}(L_0) = \boldsymbol{\omega} \mid \mathbf{V}(0) = \mathbf{a}\big). \quad\quad (54)$$

3.4 The PD - and CD policy

The analysis of IL_i for the PD - and CD policy only differs from the previous analysis in Event III. Because these simpler policies are not dependent on the state of Order 1 and Order 2, keeping track of

when these orders were placed is not necessary. Hence, defining $\mathbf{Z}_i^{(-)}$ as in (28), the analysis for the PD policy simplifies to

$$P\left(IL_i(L_0 + L_i) = m, III\right) =$$
$$\int_0^{L_0} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} P\left(D_i(\tau_z, L_0 + L_i) = R_i + u - m\right) P\left(\mathbf{V}(\tau_z) = \mathbf{z} \mid \mathbf{V}(0) = \mathbf{a}\right) \lambda_i d\tau_z, \tag{55}$$

where $u = z_0$. For the CD policy we have

$$P\left(IL_i(L_0 + L_i) = m, III\right) = \int_0^{L_0} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} p_i^{CD}(\tau_z, \mathbf{z}) P\left(\mathbf{V}(\tau_z) = \mathbf{z} \mid \mathbf{V}(0) = \mathbf{a}\right) \lambda_i d\tau_z, \tag{56}$$

where

$$p_i^{CD}(\tau_z, \mathbf{z}) = \begin{cases} P\left(D_i(\tau_z, L_0 + L_i) = R_i + Q_i - m\right) & ; \quad z_0 = Q_i \\ P\left(D_i(\tau_z, L_0 + L_i) = R_i - m\right) & ; \quad z_0 < Q_i. \end{cases} \tag{57}$$

The distribution of IL_0 is independent of the delivery policy and is obtained from (54). The expected number of units on hold, $E[O_i]$, is obviously equal to zero for the PD policy. For the CD policy there can only be between q and $Q_i - q$ units on hold at any given time, and units are only placed on hold in Event III. Therefore, defining $\mathbf{Z}_i^{(-)}(u)$ as in (43) yields

$$P\left(O_i(L_0) = u\right) = \frac{q}{\prod_{j=0}^N Q_j} \sum_{\mathbf{a} \in \mathbf{A}} \int_0^{L_0} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}(u)} P\left(\mathbf{V}(\tau_z) = \mathbf{z} \mid \mathbf{V}(0) = \mathbf{a}\right) \lambda_i d\tau_z. \tag{58}$$

The expected number of extra deliveries, S_i, is clearly equal to zero when applying the CD policy. For the PD policy, S_i is simply equal to the rate at which orders that lead to a partial delivery are placed by retailer i. Recall from Section 3.1 that partial deliveries to retailer i will occur only if retailer i places an order when $q - nQ_0 \le IL_0(L_0) \le Q_i - q - nQ_0$ ($n = 0, 1, 2\ldots$) We therefore have

$$S_i = \frac{q}{\prod_{j=0}^N Q_j} \sum_{\mathbf{a} \in \mathbf{A}} \sum_{n=0}^\infty \sum_{\boldsymbol{\omega} \in \hat{\boldsymbol{\Omega}}_i(n)} P\left(\mathbf{V}(L_0) = \boldsymbol{\omega} \mid \mathbf{V}(0) = \mathbf{a}\right) \lambda_i, \tag{59}$$

where

$$\hat{\boldsymbol{\Omega}}_i(n) = \{\boldsymbol{\omega} \in \mathbb{Z}^{N+1} \mid \omega_0 = -nQ_0 + q, -nQ_0 + 2q, \ldots, -nQ_0 + Q_i - q; \ \omega_i = R_i + 1; \\ \omega_k = R_k + 1, R_k + 2, \ldots, R_k + Q_k, \forall k \in \mathbf{N} \setminus i\}. \tag{60}$$

Optimization of reorder points

Let $TC^{MSD}(\mathbf{R})$, $TC^{PD}(\mathbf{R})$ and $TC^{CD}(\mathbf{R})$ be the total expected cost for each respective delivery policy, given a set of reorder points $\mathbf{R} = (R_0, R_1, \ldots, R_N)$. Furthermore, let $\mathbf{R}^{MSD}$, RPD and $\mathbf{R}^{CD}$ be the optimal reorder points for each policy. In this section we will utilize results from the special case of the PD

policy with $\theta_1 = \theta_2 = \ldots \theta_N = 0$. This special case has been analyzed previously (e.g. Axsäter, 2000) and we refer to it as the PD0 policy, with expected cost $TC^{PD0}(\mathbf{R})$.

We wish to minimize

$$TC^{\bullet}(\mathbf{R}) = h_0(E[IL_0^+] + \sum_{i=1}^{N} E[O_i]) + \sum_{i=1}^{N} \theta_i S_i + \sum_{i=1}^{N} (h_i E[IL_i^+] + b_i E[IL_i^-]). \tag{61}$$

We know that for a given R_0, $TC^{PD0}(\mathbf{R})$ is separable and convex in the retailer reorder points. It is obvious that the same holds true for $TC^{PD}(\mathbf{R})$. This is because the costs of the two policies are identical, except for the terms $\theta_i S_i$ which are independent of R_i ($i = 1, 2,\ldots, N$). We can therefore use the standard approach of enumerating over R_0 and, given each R_0, finding the optimal reorder points separately for each retailer (the trivial lower bound $R_i = - Q_i$ can be used as a starting point).

Lemma 6 shows that this procedure can also be applied to the CD policy.

Lemma 6

For a given value of R_0, $TC^{CD}(\mathbf{R})$ is separable and convex in R_i, $i = 1, 2,\ldots, N$.

Proof

As stated previously, $TC^{CD}(\mathbf{R})$ is separable because the choice of reorder point only affects the retailer in question. To prove convexity, we show that the retailer cost function is a sum of convex functions. First note that the distributions of O_i and IL_0 are independent of R_i. Defining X_i so that $IP_i = R_i + X_i$ means that $X_i(L_0) - B_i(L_0) - O_i(L_0)$ is independent of R_i, as well as independent of the demand after time L_0. From (8) we have that $IL_i(L_0+L_i) = R_i + X_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0,L_0+L_i)$. Given $X_i(L_0) - B_i(L_0) - O_i(L_0) = \alpha$, retailer costs are convex in R_i as the analysis is equivalent to a base-stock single-echelon system (with base-stock level $R_i + \alpha$). Summation of these convex functions for all possible values of α, multiplied with their corresponding probabilities, yields the retailer cost function under the CD policy. ■

It can be shown by examples that $TC^{MSD}(\mathbf{R})$ is not always convex in R_i for a given R_0. In fact, in some cases it has multiple local minima. However, we can obtain bounds for the optimal retailer reorder points. Let $R_i^{\bullet}(R_0)$ denote the optimal reorder point given R_0 for a specific delivery policy (in case the optimal solution is not unique, let $R_i^{PD}(R_0)$ be the smallest - and $R_i^{CD}(R_0)$ be the largest reorder point that is optimal). We have the following bounds:

Lemma 7

$R_i^{PD}(R_0) \leq R_i^{MSD}(R_0) \leq R_i^{CD}(R_0)$ for $i = 1, 2,\ldots, N$.

Proof

Given the MSD policy, assume that the system is in a state such that a placement of an order by retailer i requires an MSD decision with given values of the parameters t_1, t_2 and u. Let the rate at which such orders are placed be $\mu_i(t_1,t_2,u)$. Note that $\mu_i(t_1,t_2,u)$ is independent of R_j (j = 1, 2,..., N). The difference between the PD policy and the MSD policy is that an opportunity for cost savings occurs each time the MSD decision is made. The amount saved by the MSD decision compared to partial delivery is $(\theta_i - \Delta C_i(R_i,t_1,t_2,u), 0)^+$ (see Proposition 1). Thus, the expected total cost for the MSD policy can be obtained by subtracting the expected cost savings of the MSD decisions from the expected total cost of the PD policy. That is, for any given reorder points we have that

$$TC^{MSD}(\mathbf{R}) = TC^{PD}(\mathbf{R}) - E_{u,t_1,t_2}\left[\sum_{j=1}^{N}\mu_j(u,t_1,t_2)\left(\theta_j - \Delta C_j(R_j,t_1,t_2,u),0\right)^+\right]. \tag{62}$$

Now assume that $R_i^{MSD}(R_0) < R_i^{PD}(R_0)$. This implies that

$$TC^{PD}(R_0,R_1,\ldots,R_i^{MSD}(R_0),\ldots,R_N) > TC^{PD}(R_0,R_1,\ldots,R_i^{PD}(R_0),\ldots,R_N). \tag{63}$$

Furthermore, if $\Delta C_i(R_i,t_1,t_2,u)$ is decreasing in R_i,

$$\left(\theta_i - \Delta C_i(R_i^{MSD},t_1,t_2,u),0\right)^+ \geq \left(\theta_i - \Delta C_i(R_i^{PD},t_1,t_2,u),0\right)^+, \tag{64}$$

and therefore from (62) we have that

$$TC^{MSD}(R_0,R_1,\ldots,R_i^{MSD}(R_0),\ldots,R_N) > TC^{MSD}(R_0,R_1,\ldots,R_i^{PD}(R_0),\ldots,R_N). \tag{65}$$

However, (65) is a contradiction and thus it must hold that $R_i^{PD}(R_0) \leq R_i^{MSD}(R_0)$, if $\Delta C_i(R_i,t_1,t_2,u)$ is decreasing in R_i. Analogous reasoning can be applied for the CD policy. In this case

$$TC^{MSD}(\mathbf{R}) = TC^{CD}(\mathbf{R}) - E_{u,t_1,t_2}\left[\sum_{j=1}^{N}\mu_j(u,t_1,t_2)\left(\Delta C_j(R_j,t_1,t_2,u) - \theta_j,0\right)^+\right], \tag{66}$$

and it follows that $R_i^{MSD}(R_0) \leq R_i^{CD}(R_0)$. It remains to show that $\Delta C_i(R_i,t_1,t_2,u)$ is decreasing in R_i. Manipulating the expressions in (2), (3) and (4) yields

$$\Delta C(R_i+1,t_1,t_2,u) - \Delta C(R_i,t_1,t_2,u) = \sum_{n=1}^{u}\Delta g(R_i,n), \text{ where}$$

$$\Delta g(R_i,n) = \begin{cases} \dfrac{h_i+b_i}{\lambda_i}\left(\displaystyle\sum_{j=0}^{R_i+n}\dfrac{(\lambda_i(t_2+L_i))^j}{j!}e^{-\lambda_i(t_2+L_i)} - \sum_{j=0}^{R_i+n}\dfrac{(\lambda_i(t_1+L_i))^j}{j!}e^{-\lambda_i(t_1+L_i)}\right); & R_i+n \geq 0 \\ 0 & ; R_i+n < 0. \end{cases} \tag{67}$$

We identify the two sums in (67) as the cumulative distribution functions for Poisson variables with means $\lambda_i(t_2 + L_i)$ and $\lambda_i(t_1 + L_i)$, respectively. Because $t_1 < t_2$, the latter mean is lower and therefore the value of the associated cumulative distribution function evaluated at $R_i + n$ is higher. Hence, the difference in (67) is less than or equal to zero, meaning that $\Delta C_i(R_i,t_1,t_2,u)$ is decreasing in R_i. ∎

To summarize, our optimization procedure means that we increase R_0 by one unit at a time, starting with $R_0 = 0$. In each step we determine $R_i^{\bullet}(R_0)$, $i = 1, 2, \ldots, N$. (by using the lower bounds and convexity for the PD - and CD policy, and by considering all values that are given by the lower and upper bound for the MSD policy). Because IL_0^+ is increasing in R_0, we can stop increasing R_0 when $h_0 IL_0^+(R_0)$ is larger than the lowest expected total cost found so far.

4. Numerical experiments

To demonstrate the performance of the different delivery policies, we consider 32 problem scenarios. All scenarios feature three retailers with $L_0 = 4$, $h_0 = h_i = 1$ ($\forall i$), and the order quantities set to two different levels; either $Q_0 = 4$, $Q_1 = 2$. $Q_2 = 3$, $Q_3 = 4$, or $Q_0 = 6$, $Q_1 = 4$. $Q_2 = 5$, $Q_3 = 6$. The remaining input parameters are identical between retailers. We have $\theta_i = \{2, 4, 8, 16\}$, $b_i = \{5, 50\}$ and $L_i = \{1, 2\}$ ($\forall i$), where all combinations of the parameter levels constitute our problem set.

The MSD policy will always produce the lowest expected total cost. Hence, it is natural to use it as a base for the comparison between the different delivery policies. Let ΔP and ΔCD denote the relative cost increase of the PD - and CD policy compared to the MSD policy. That is,

$$
\begin{aligned}
\Delta PD &= \frac{TC^{PD} - TC^{MSD}}{TC^{MSD}} \\
\Delta CD &= \frac{TC^{CD} - TC^{MSD}}{TC^{MSD}}.
\end{aligned}
\tag{68}
$$

Table 1 provides the input data, the optimal solutions and the corresponding expected costs for each of the three delivery policies. In Table 1 we see that the PD and CD policy perform significantly worse than the MSD policy in many scenarios. The maximum ΔPD was 26.6% and the maximum ΔCD was 17.9%, with averages of 5.8% and 5.9%, respectively. This indicates that there can be a significant advantage in using our new MSD policy. One can also compare the MSD policy to the better of the two other policies (that is, to consider min$\{\Delta PD, \Delta CD\}$). This isolates the specific cost increase of not using state-dependent deliveries, and the results show a maximum increase of 6.1% (the average was 1.8%). Table 1 shows that the maximum value occurs when the handling cost is at an intermediate value ($\theta_i = 8$). This is logical because the PD - and CD policy will obviously perform well when the handling cost is low - and high, respectively. Figure 5 depicts the average ΔPD - and ΔCD values for given values of the handling cost.

Table 1. Input data and results of numerical tests.

|	Input data				Results MSD		Results PD		Results CD		Comparison	
#	θ_i	b_i	L_i	**Q**	$\mathbf{R^{MSD}}$	TC^{MSD}	$\mathbf{R^{PD}}$	TC^{PD}	$\mathbf{R^{CD}}$	TC^{CD}	ΔPD	ΔCD
1	2	5	1	(4,2,3,4)	(10,1,1,1)	10.02	(10,1,1,1)	10.18	(10,1,1,1)	10.70	1.6%	6.8%
2	2	5	1	(6,4,5,6)	(8,1,1,0)	11.91	(8,1,1,0)	11.98	(9,1,1,1)	14.04	0.6%	17.9%
3	2	5	2	(4,2,3,4)	(9,3,2,2)	11.28	(9,3,2,2)	11.44	(10,3,2,2)	11.98	1.4%	6.2%
4	2	5	2	(6,4,5,6)	(6,3,2,2)	12.99	(6,3,2,2)	13.06	(10,2,2,2)	15.06	0.5%	15.9%
5	2	50	1	(4,2,3,4)	(12,3,3,2)	16.26	(12,3,3,2)	16.37	(12,3,3,3)	17.36	0.7%	6.8%
6	2	50	1	(6,4,5,6)	(11,3,2,2)	18.79	(11,3,2,2)	18.84	(14,2,2,2)	20.95	0.3%	11.5%
7	2	50	2	(4,2,3,4)	(11,5,5,4)	18.47	(12,5,4,4)	18.62	(11,5,5,5)	19.47	0.8%	5.4%
8	2	50	2	(6,4,5,6)	(10,5,4,4)	20.88	(11,4,4,4)	20.93	(14,4,4,4)	23.01	0.3%	10.2%
9	4	5	1	(4,2,3,4)	(10,1,1,1)	10.31	(10,1,1,1)	10.82	(10,1,1,1)	10.70	5.0%	3.8%
10	4	5	1	(6,4,5,6)	(8,1,1,0)	12.35	(8,1,1,0)	12.59	(9,1,1,1)	14.04	2.0%	13.7%
11	4	5	2	(4,2,3,4)	(10,3,2,2)	11.60	(10,3,2,2)	12.15	(10,3,2,2)	11.98	4.8%	3.3%
12	4	5	2	(6,4,5,6)	(7,3,2,2)	13.45	(7,3,2,2)	13.73	(10,2,2,2)	15.06	2.1%	11.9%
13	4	50	1	(4,2,3,4)	(12,3,3,2)	16.51	(13,3,2,2)	16.76	(12,3,3,3)	17.36	1.5%	5.1%
14	4	50	1	(6,4,5,6)	(11,3,2,2)	19.14	(11,3,2,2)	19.30	(14,2,2,2)	20.95	0.8%	9.4%
15	4	50	2	(4,2,3,4)	(12,5,4,4)	18.74	(12,5,4,4)	19.08	(11,5,5,5)	19.47	1.8%	3.8%
16	4	50	2	(6,4,5,6)	(11,4,4,4)	21.23	(11,4,4,4)	21.39	(14,4,4,4)	23.01	0.8%	8.4%
17	8	5	1	(4,2,3,4)	(10,1,1,1)	10.57	(12,1,1,0)	11.94	(10,1,1,1)	10.70	12.9%	1.2%
18	8	5	1	(6,4,5,6)	(9,1,1,0)	13.00	(10,1,0,0)	13.78	(9,1,1,1)	14.04	6.0%	8.0%
19	8	5	2	(4,2,3,4)	(10,3,2,2)	11.85	(11,2,2,2)	13.44	(10,3,2,2)	11.98	13.5%	1.1%
20	8	5	2	(6,4,5,6)	(8,2,2,2)	14.12	(8,2,2,2)	14.97	(10,2,2,2)	15.06	6.1%	6.6%
21	8	50	1	(4,2,3,4)	(13,3,2,2)	16.84	(13,3,2,2)	17.51	(12,3,3,3)	17.36	3.9%	3.0%
22	8	50	1	(6,4,5,6)	(12,3,2,2)	19.67	(13,2,2,2)	20.06	(14,2,2,2)	20.95	2.0%	6.5%
23	8	50	2	(4,2,3,4)	(12,5,4,4)	19.06	(13,4,4,4)	19.98	(11,5,5,5)	19.47	4.8%	2.1%
24	8	50	2	(6,4,5,6)	(11,4,4,4)	21.81	(12,4,4,3)	22.25	(14,4,4,4)	23.01	2.0%	5.5%
25	16	5	1	(4,2,3,4)	(10,1,1,1)	10.69	(14,1,0,0)	13.53	(10,1,1,1)	10.70	26.6%	0.1%
26	16	5	1	(6,4,5,6)	(9,1,1,0)	13.64	(12,0,0,0)	15.52	(9,1,1,1)	14.04	13.8%	3.0%
27	16	5	2	(4,2,3,4)	(10,3,2,2)	11.97	(13,2,2,1)	15.14	(10,3,2,2)	11.98	26.5%	0.1%
28	16	5	2	(6,4,5,6)	(9,2,2,2)	14.71	(12,2,1,1)	16.90	(10,2,2,2)	15.06	14.9%	2.3%
29	16	50	1	(4,2,3,4)	(13,3,3,2)	17.15	(15,2,2,2)	18.54	(12,3,3,3)	17.36	8.1%	1.2%
30	16	50	1	(6,4,5,6)	(13,2,2,2)	20.28	(14,2,2,2)	21.26	(14,2,2,2)	20.95	4.9%	3.3%
31	16	50	2	(4,2,3,4)	(12,5,5,4)	19.34	(14,4,4,4)	21.25	(11,5,5,5)	19.47	9.9%	0.6%
32	16	50	2	(6,4,5,6)	(12,4,4,4)	22.47	(13,4,4,3)	23.64	(14,4,4,4)	23.01	5.2%	2.4%
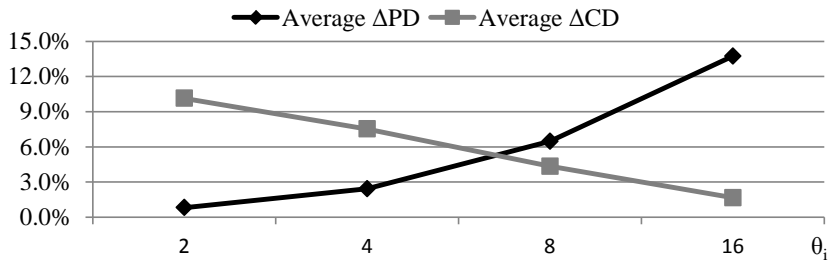
Figure 5. Average ΔPD - and ΔCD values for given values of the handling cost.

Figure 5 illustrates how performance of the PD policy deteriorates as the cost of partial deliveries increases. It is easy to show that the cost increase of using the PD policy compared to the MSD policy is unbounded as the handling cost increases towards infinity. However, the cost increase of using the CD policy is bounded. For this policy, the cost increase is at its highest value when the handling cost is zero, and then it decreases to zero (recall that Corollary 2 implies that the CD and MSD policies are equivalent above a certain value of the handling cost).

As mentioned in Section 1, the previous literature has almost exclusively considered systems with partial delivery policies. In these systems the optimal solutions usually point toward having a small proportion of stock at the central warehouse, thus having the main part at the retailers. This means that the retailers keep most of the safety stock, and the central warehouse provides relatively low service (e.g., Axsäter, 2003). To investigate how these results carry over to our new policies, Figure 6 shows how the optimal proportion of stock under the MSD policy changes with the handling cost. For each value of the handling cost there are eight different scenarios. Hence, the proportion of stock is defined as the sum of the expected positive central warehouse inventory over the eight scenarios, divided by the sum of the expected positive retailer inventories over the eight scenarios. Note that, in addition to the scenarios in Table 1, Figure 6 also includes results from the same eight scenarios with the handling cost equal to zero (because $h_0 = h_i$, $\forall i$, in this case the MSD policy is equal to the PD0 policy considered in previous work). Moreover, Figure 6 also includes the proportion with the handling cost equal to the upper bound (UB) provided by Corollary 2 (where the MSD policy is equal to the CD policy).
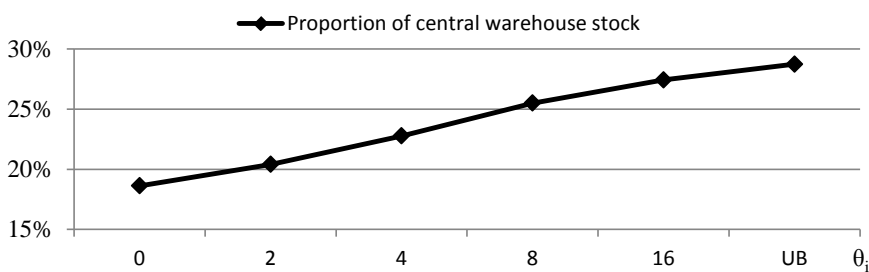

Figure 6. Proportion of total expected stock stored at the central warehouse for given values of the handling cost under the MSD policy.

We see in Figure 6 that increasing the handling cost makes it more attractive to keep stock at the central warehouse. This seems natural because the handling cost penalizes situations where there is not enough stock on hand to cover an entire order of Q_i. Hence, it appears that under our more general cost structure, more stock will be allocated to the central warehouse than seen previously in the

literature. However, even for high handling costs where the CD policy is optimal, the majority of stock is still kept at the retailers.

5. Summary and concluding remarks

We have considered a two-echelon inventory model consisting of one central warehouse and a number of retailers. The purpose has been to evaluate the cost impact of different central warehouse delivery policies. This is done by introducing a more realistic cost structure for the handling of orders than previously considered in the literature. We have provided an exact method for cost evaluation and optimization of the reorder points under three different delivery policies: the partial delivery policy, the complete delivery policy and the mixed state-dependent policy. The state-dependent policy makes a cost minimizing decision between partial or complete deliveries for each retailer order and, thus, gives a performance guarantee compared to the simpler policies.

Our results show that the choice of delivery policy can have a significant impact on the operating costs of the considered system. Analytically, we have shown that the mixed state-dependent policy never performs worse than the two other policies. We have also identified sufficient conditions for when one should always choose complete deliveries. Numerically, when using our new state-dependent policy as a benchmark, we recorded maximum relative cost increases of 26.6% (average 5.8%) for the partial delivery policy, and 17.9% (average 5.9%) for the complete delivery policy. Hence, analytical as well as numerical results show that the common assumption of partial deliveries may be questionable in many cases. Moreover, our results suggest that, under our new cost structure, it is optimal to allocate more stock to the central warehouse than recorded previously in the literature.

Finally, we believe that the method of analysis that is presented in this work holds a high degree of generality. For instance, by deriving the distributions of the inventory levels it is easy to consider various types of service constraints, as opposed to backorder costs. By using our approach based on the nominal inventory position it could, for example, also be possible to consider more general demand distributions such as compound Poisson.

References

Andersson, J. 1999. Coordinated multi-stage inventory systems with stochastic demand. Doctoral Thesis, Lund University.

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

Axsäter, S. 1993. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* **41** 777-785.

Axsäter, S. 1997. Simple evaluation of echelon stock (R, Q)-policies for two-level inventory systems. *IIE Transactions* **29** 661–669.

Axsäter, S. 1998. Evaluation of installation stock based (R,Q) policies for two-level inventory systems with Poisson demand. *Operation Research* **46** S135-S145.

Axsäter, S. 2000. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand. *Operations Research* **48** 686-696.

Axsäter, S. 2003. Supply chain operations: Serial and distribution inventory systems. *Handbooks in operations research and management science* **11**. S.C. Graves, A.G. de Kok (editors). Elsevier, Amsterdam, The Netherlands, 525-559.

Axsäter, S. 2007. On the first come-first served rule in multi-echelon inventory control. *Naval Research Logistics* **54** 485-491.

Axsäter, S., J. Marklund. 2008. Optimal "position based" warehouse ordering in divergent two-echelon inventory systems. *Operations Research* **56** 976-991.

Chen, F., Y.S. Zheng. 1997. One warehouse multi-retailer systems with centralized stock information. *Operations Research* **45** 275-287.

Chu, L.Y., Z.M. Shen. 2010. A power-of-two ordering policy for one warehouse multi-retailer systems with stochastic demand. *Operations Research* **58** 492-502.

Forsberg, R. 1995. Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research* **81** 143-153.

Forsberg, R. 1997. Exact evaluation of (R,Q)-policies for two-level inventory systems with Poisson demand. *European Journal of Operational Research* **96** 130-138.

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

Graves, S.C. 1996. A multi-echelon inventory model with fixed replenishment intervals. *Management Science* **42** 1-18.

Howard, C., J. Marklund. 2011. Evaluation of stock allocation policies in a divergent inventory system with shipment consolidation. *European Journal of Operational Research* **211** 298-309.

Howard, C. 2013. New allocation policies for divergent inventory systems with real-time information and shipment consolidation. Working paper Lund University.

Marklund, J. 2002. Centralized inventory control in a two-level distribution system with Poisson demand. *Naval Research Logistics* **49** 798-822.

Marklund, J. 2006. Controlling inventories in divergent supply chains with advance-order information. *Operations Research* **54** 988-1010.

Marklund, J. 2011. Inventory control in divergent supply chains with time-based dispatching and shipment consolidation. *Naval Research Logistics* **58** 59-71.

Marklund, J., K. Rosling. 2012. Lower bounds and heuristics for supply chain stock allocation. *Operations Research* **60** 92-105.

Stenius, O., G. Karaarslan, J. Marklund, A.G. de Kok. 2013. Exact analysis of divergent inventory systems with time-based shipment consolidation and compound Poisson demand. Working paper Lund University and Eindhoven University of Technology.

Paper IV

Divergent Two-echelon Inventory Systems with Compound Renewal Demand

Olof Stenius

Department of Industrial Management and Logistics, Lund University
olle.stenius@iml.lth.se

Abstract

This paper studies a two-echelon inventory system consisting of one warehouse and multiple non-identical retailers facing independent compound renewal demand. There are linear holding costs at all stock points and backorder costs at the retailers. All stock points apply continuous base stock replenishment policies, first-come-first-served allocations, partial order fulfillments and complete backordering. For this system, the exact long run average inventory level distribution is determined. Through this distribution, optimal order-up-to levels of all stock points are obtained using a recursive cost evaluation procedure. Numerical tests show that it can be very costly to assume exponential inter-arrival times, as is customary in many existing models.

1. Introduction

Developments in Information Technologies enable firms to collect more detailed information about their businesses. For logistics planning, this means that many companies have access to detailed information about their customer demand, for instance the distribution of the time between consecutive customer arrivals (the inter-arrival time) and demand sizes. In order to benefit from this information, models that can handle general demand distributions are needed. Existing exact inventory control models of distribution systems predominantly assume exponential inter-arrival times of customers. The memoryless property of the resulting Poisson processes has many analytical advantages. In many cases, especially when there are a large number of independent customers arriving to each stock point, this is also a well-motivated assumption. However, in other situations, for example for critical spare parts, it is not uncommon to have a local stock point (or consignment stock) in conjunction to larger customers. In these systems, e.g. increasing or decreasing underlying failure rates or preventive maintenance can make the inter-arrival times far from exponential. In other systems, customers' batching of orders makes exponential times poor estimates of reality.

In this context, we present a one-warehouse-multiple-retailer inventory model, where the inter-arrival times of customer demands follow any continuous distribution and the customer demand sizes are positive integers, at each retailer. The distributions of the inter-arrival times and demand sizes at all retailers are known. Assuming that consecutive inter-arrival times and demand sizes are independent (also across the retailers), the customer demands at the retailers correspond to independent compound renewal processes.

The model includes linear holding costs per unit and time unit at all stock points and backorder costs per unit and time unit at the retailers. The retailers replenish from the central warehouse, which, in turn, replenish from an outside supplier. All stock points apply continuous base stock policies to replenish their inventory. The optimal replenishment policy for this type of distribution system is unknown even for simpler demand distributions. However, for systems with negligible replenishment costs compared to the holding and backorder costs (for instance expensive spare parts) continuous base stock (or order-up-to S) policies are optimal in many single-echelon systems and also commonly used in multi-echelon systems both in theory and in practice. First-Come-First-Served (FCFS) allocations are assumed at all stock points. This is motivated by tractability reasons and their wide spread use. There are also several studies suggesting close to optimal performance in many distribution systems (see Graves, 1996, Howard and Marklund, 2011, and Howard, 2013).

The main contribution is the exact analysis and derivation of long run inventory level distributions at all stock points. Through these distributions the expected long run system cost is obtained and a recursive optimization procedure for the order-up-to-levels at all stock points is provided. Our numerical study show that the common assumption of exponential inter-arrival times can be extremely costly (cost increases of more than 200%).

The rest of this section is dedicated to a literature review and Section 2 formulates the model. Section 3 presents the analysis of the inventory level distributions and costs of the system. Based on this analysis, Section 4 shows how to optimize the order-up-to levels with a recursive procedure. Section 5 presents the numerical study, and Section 6 concludes.

Related literature

As mentioned before, thus far, exact evaluation techniques for one-warehouse-multiple-retailer continuous review inventory control systems have been based on Poisson processes. Simon (1971) presents an exact method for evaluating the expected inventory levels and thereby the costs in a system were customer demand is Poisson distributed and all stock points replenish with order-up-to S policies. For this system, Axsäter (1990) provides a fast recursive approach for determining the costs. The Axsäter model is generalized to compound Poisson demand by Forsberg (1995) and to installation stock (R,Q)-policies and Poisson demand by Axsäter (1993a) and Forsberg (1997a). Forsberg (1997b) extends the analysis of Forsberg (1997a) to the case were customer inter-arrival rates are k-Erlang distributed. Axsäter (2000) is related to our work as it also evaluates the inventory level distributions under compound Poisson and installation stock (R,Q) policies. This is also the case for Chen and Zheng (1997) that analyses echelon stock (R,Q) policies exactly for Poisson and as an approximation for compound Poisson. For an overview of more recent exact models, that analyzes more elaborate replenishment policies in distribution systems; see, for example, Marklund (2011).

To the best of our knowledge, Forsberg (1997b) provides the only exact analysis of costs with non-Poisson customer inter-arrival times in a continuous review multi-echelon distribution system. This analysis is also based on the Poisson process as it uses an adjusted Poisson process with a higher *virtual* demand frequency and assumes that only every k *virtual* demands trigger orders and generates costs. Compared to Forsberg (1997b) our present work distinguishes itself in several ways; (i) we consider general continuous distributions of the inter-arrival times, (ii) we consider a compounding distribution, i.e. the customers can order multiple units at once, (iii) we evaluate the expected inventory level distribution, not only the costs, (iv) we provide an optimization procedure for the order-up-to levels, and (v) we present a new cost evaluation technique.

Among the existing approximation models for continuous review inventory control of distribution systems Kiesmüller et al. (2004) and Forsberg (1997b) are of specific interest to our research, as they approximate general (compound) renewal demand. Our research is also related to the large body of research examining periodic review inventory distribution systems. Particularly Graves (1996), Axsäter (1993b), and Shang, Tao and Zhou (2015) are related to the present work as they assume FCFS allocation policies in their model. Also, Erkip, Hausman and Nahmias (1990) relates to our work as they study a system where the demands in successive time periods are dependent. When the customer arrivals follow general renewal processes, as they do in our present work, the demand in consecutive time periods (of constant length) is usually dependent, as opposed to models assuming Poisson arrivals.

General renewal or compound renewal demand has also been considered in the single-echelon settings, see, for instance, Kruse (1981), Larsen and Thorstenson (2008), Syntetos et al. (2015) and references therein.

2. Problem formulation

As mentioned, we study a 2-echelon system where a central warehouse supplies N non-identical retailers, which face independent compound renewal demand. The retailers replenish from the central warehouse (no direct demand at the warehouse or lateral transshipments are allowed) and the warehouse replenish from an outside supplier. The transportation time to the retailers including picking, loading, transporting, receiving etc. is constant, as is the replenishment lead time, L_0, for the central warehouse (the time from placement of an order with the outside supplier until the units are available at the warehouse). Alternatively the model describes a repairable item system, where the central warehouse repairs all broken parts in L_0 time units.

Recall that, there are linear holding costs at all stock points (per unit and time unit) and linear backorder costs at all retailers, and all stock points apply order-up-to S policies and FCFS allocations. When orders cannot be fully satisfied, all stock points use complete backordering and partial order fulfillment, which is common for expensive spare parts (where shipment costs are negligible). The order-up-to levels are assumed to be non-negative at all stock points. For the retailers, this is optimal (given that base stock policies are used), but for the warehouse, this is not necessarily the case. There exist systems where negative order-up-to levels can be optimal at the central warehouse (especially for some systems where the inter-arrival times are close to deterministic). The restriction is motivated by analytical tractability, the fact that it is optimal for exponential inter-arrival times, and its use in practice. We define:

Ω	=	set of retailers (number of retailers = N)
S_0	=	order up to level at the central warehouse
S_i	=	order up to level at retailer i
L_0	=	replenishment lead time to the central warehouse
L_i	=	transportation time from the central warehouse to retailer i
h_0	=	holding cost per unit and time unit at the central warehouse
h_i	=	holding cost per unit and time unit at retailer i
b_i	=	backorder cost per unit and time unit at retailer i
A_i	=	inter-arrival time of customers at retailer i
$f_X(\tau)$	=	probability density function of the continuous variable X
$F_X(\tau)$	=	cumulative distribution function of the continuous variable X
$q_i(x)$	=	probability that the customer demand size at retailer i is equal to x
$q_i'(x)$	=	probability that the customer demand size at retailer i is larger than x
λ_i	=	long run average number of customers arriving per time unit at retailer i = $1/E[A_i]$
μ_i	=	average order size at retailer i
$IL_i(t)$	=	inventory level (= stock on hand – backorders) at retailer i at time t
$IL_0(t)$	=	inventory level at central warehouse at time t
x^+	=	max(x,0), $x^- = \max(-x,0)$

3. Cost Analysis

In this section the long run average costs per time unit is determined by a steady state analysis. Because the inter-arrival times are stochastic with a continuous distribution (non-lattice) and independent, the system will eventually reach steady state. The costs for the time period until steady state is reached are finite, and will therefore not affect the long run average costs. All proofs are deferred to Appendix A. We define:

$TC(S_0,S_1,\ldots,S_N)$ = the expected total costs per time unit for the system in steady state, with order-up-to levels S_0 and S_i $\forall i \in \Omega$.

$C_0(S_0)$ = the expected costs per time unit in steady state at the central warehouse, when the order-up-to level at the warehouse is S_0

$C_i(S_0,S_i)$ = the expected costs per time unit in steady state at retailer i, when the order-up-to level at the warehouse is S_0 and the order-up-to level at retailer i is S_i

4

$D_i[t_1,t_2) =$ total customer demand at retailer i in time interval $[t_1,t_2)$

$D_i(t_1,t_2) =$ total customer demand at retailer i in time interval (t_1,t_2)

$\hat{\pi}_i^T(x) =$ probability of a demand of x units at retailer i during a time interval of T time units excluding the initial demand that marks the start of this interval

$\underline{\pi}_i^T(x) =$ probability of a demand of x units at retailer i during a time interval of T time units excluding the demand that marks the end of this interval

$\pi_i^T(x) =$ probability of a demand of x units at retailer i during T time units in steady state

$\pi_i^{[T_1,T_2]}(x,y) =$ joint probability of a demand of x and y units at retailer i during two consecutive time intervals of T_1 and T_2 time units each in steady state

$\pi_\Theta^T(x) =$ probability of a total demand of x units for the retailer set Θ during a time interval of T time units in steady state

How to determine these $\pi_.^.(\bullet)$-probabilities is a key feature of the cost analysis, and it is described in Section 3.3 below.

The expected total costs in the system per time unit in steady state is

$$TC(S_0,S_1,...,S_N) = C_0(S_0) + \sum_{i\in\Omega} C_i(S_0,S_i). \tag{1}$$

The costs at the central warehouse, $C_0(S_0)$, consist of only holding costs. The analysis is based on studying the distribution of the inventory level at time t in steady state. We have

$$\begin{aligned}C_0(S_0) &= h_0 E\left[IL_0^+(t)\right] \\ &= h_0 \sum_{m=1}^{S_0} mP(IL_0(t) = m).\end{aligned} \tag{2}$$

The inventory position at the central warehouse (= outstanding orders + stock on hand – backorders) is by definition equal to the base stock level S_0. Because of the continuous review base stock policy at each retailer i, every customer demand immediately creates a retailer order of the same size. Using the inventory balance equation, we get

$$IL_0(t) = S_0 - \sum_{i\in\Omega} D_i[t-L_0,t). \tag{3}$$

In steady state, the long run probability of an inventory level m at the central warehouse can thus be obtained as

$$P(IL_0(t) = m) = \pi_\Omega^{L_0}(S_0 - m) \tag{4}$$

How to determine the probability $\pi_\Omega^{L_0}(S_0 - m)$ in (4) is described in Section 3.3.

5

The retailer costs, $C_i(S_0,S_i) \ \forall \, i$, consist of both holding and backorder costs. For analyzing these costs, we study the inventory level at retailer i at time t in steady state. The solution is obtained by analyzing the system behavior during the $L_0 + L_i$ time units prior to t. For ease of exposition, we let $t = L_0 + L_i$. We also define the amount of backordered units at the central warehouse destined for retailer i at time $t - L_i = L_0$, $B_i(S_0)$. The inventory level at retailer i at time $L_0 + L_i$, $IL_i(L_0 + L_i)$, is

$$IL_i(L_0 + L_i) = S_i - B_i(S_0) - D_i[L_0, L_0 + L_i]. \tag{5}$$

The analysis of the inventory levels at the retailers is explained in Section 3.1 below. Based on the inventory level distribution, the expected costs at retailer i is obtained as

$$\begin{aligned}
C_i(S_0,S_i) &= h_i E\left[IL_i^+(L_0 + L_i)\right] + b_i E\left[IL_i^-(L_0 + L_i)\right] \\
&= (h_i + b_i)\sum_{m=1}^{S_i} mP\left(IL_i(L_0 + L_i) = m\right) - b_i\left(S_i - E\left[B_i(S_0)\right] - \lambda_i \mu_i L_i\right).
\end{aligned} \tag{6}$$

(6) follows as for any variable X, $E[X] = E[X^+] - E[X^-]$ and the expected demand between L_0 and $L_0 + L_i$ is $\lambda_i \mu_i L_i$. $E[B_i(S_0)]$ is analyzed in Section 3.2.

3.1. Inventory Level at Retailer i

The derivation of the inventory level probabilities at the retailers is more challenging than the analysis of the inventory level at the central warehouse. The reason is that the warehouse is usually unable to fulfill all retailer orders immediately. The difficulties in analyzing (5) for a specific retailer i, lies in determining $B_i(S_0)$, keeping in mind that, because of the compound renewal demand, $B_i(S_0)$ (at time L_0) can be dependent on $D_i(L_0,L_0 + L_i)$. For analyzing $B_i(S_0)$ we define the nominal inventory position at time τ to be

$$\Psi(\tau) = S_0 - \sum_{i\in\Omega} D_i[0,\tau], \qquad \text{for } 0 \le \tau \le L_0. \tag{7}$$

The nominal inventory position is a stepwise decreasing function. When positive, it provides information about how many units (not yet demanded by the retailers) the central warehouse still can provide (to any retailer) before time L_0. Lemma 1 summarizes some of the properties of the nominal inventory position, illustrated in Figure 1. Similar definitions of the nominal inventory position and its properties have previously been used in Stenius et al. (2015) and Howard and Stenius (2014) for other types of distribution systems under more restrictive Poisson/compound Poisson demand.

Lemma 1. *Properties of the nominal inventory position,* $\Psi(\tau)$, $0 \le \tau \le L_0$:

a) $\Psi(0) = S_0$

b) $\Psi(L_0) = IL_0(L_0)$

c) if $\Psi(\tau) \geq 0$, *no retailer orders before time* τ *will be backordered at the central warehouse at time* L_0

d) if $\Psi(\tau) \leq 0$, *all retailer orders in time interval* (τ, L_0) *will be backordered at the central warehouse at time* L_0

e) if $\Psi(\tau) = x$, $x \geq 0$ *when retailer* i *places an order of* $y > x$ *units,* x *of these units will be delivered before time* L_0 *and* y-x *of these units will be backordered at time* L_0.
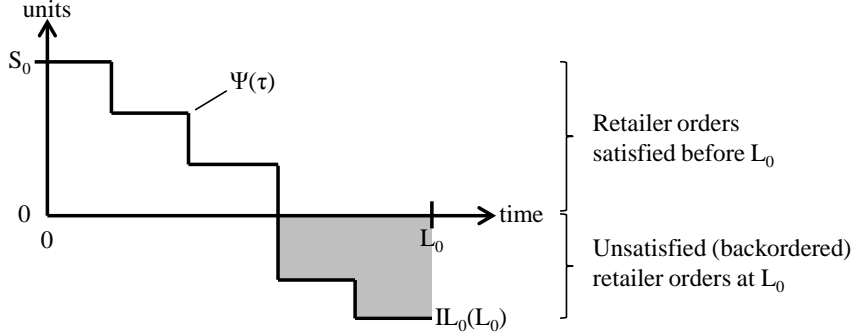

Figure 1. Sample path of the nominal inventory position from time 0 to L_0.

Based on the properties of the nominal inventory position, we divide the analysis of the inventory level at the considered retailer i at time $L_0 + L_i$ in three events:

E1: $IL_0(L_0) \geq 0$

E2: At time τ $(0 \leq \tau < L_0)$ an order from retailer $j \neq i$ occurs that brings the nominal inventory position from a non-negative to a negative value.

E3: At time τ $(0 \leq \tau < L_0)$ an order from retailer i occurs that brings the nominal inventory position from a non-negative to a negative value.

These events are mutually exclusive and collectively exhaustive and the probability of an inventory level m at retailer i at time $L_0 + L_i$ in steady state is thus

$$
\begin{aligned}
&P\big(IL_i\big(L_0 + L_i\big) = m\big) \\
&= P\big(IL_i\big(L_0 + L_i\big) = m, E1\big) + P\big(IL_i\big(L_0 + L_i\big) = m, E2\big) + P\big(IL_i\big(L_0 + L_i\big) = m, E3\big).
\end{aligned}
\tag{8}
$$

The probability of an inventory level m at retailer i for event E1

For event E1, the inventory level at the central warehouse is non-negative at time L_0. For determining the probability mass functions of the inventory levels at retailer i in this event we use Lemma 2 and Corollary 1.

Lemma 2. *For* $IL_0(L_0) \geq 0$ *(event E1)*

$$IL_i\left(L_0 + L_i\right) = S_i - D_i\left[L_0, L_0 + L_i\right) \tag{9}$$

Corollary 1. $IL_i(L_0 + L_i) = m$ *for event E1 iff conditions a), b) and c) are satisfied:*

a) $\sum_{j \in \Omega \setminus [i]} D_j\left[0, L_0\right) = x$ $\qquad\qquad\qquad , x \le S_0$

b) $D_i\left[0, L_0\right) = y$ $\qquad\qquad\qquad\qquad , y \le S_0 - x$

c) $D_i\left[L_0, L_0 + L_i\right) = S_i - m$

Note that because of the compound renewal demand processes at the retailers, the demands in consecutive time intervals at the same retailer are usually dependent. For that reason, conditions b) and c) in Corollary 1 are dependent. They are however independent of a) as demand at different retailers are independent. Because we have no specific requirements on when an order should occur, the probability of condition a) for a given x is $\pi_{\Omega \setminus [i]}^{L_0}(x)$. The joint probability of conditions b) and c) for a given y is $\pi_i^{[L_0, L_i]}(y, S_i - m)$. How to determine these $\pi_{\cdot}^{\cdot}(\bullet)$-probabilities is explained in Section 3.3. Considering all possible values of x and y, the probability of an inventory level m at retailer i for event E1 at an arbitrary point in time in steady state (at time $L_0 + L_i$) can be obtained as

$$P\left(IL_i\left(L_0 + L_i\right) = m, E1\right) = \sum_{x=0}^{S_0} \pi_{\Omega \setminus [i]}^{L_0}(x) \sum_{y=0}^{S_0 - x} \pi_i^{[L_0, L_i]}(y, S_i - m). \tag{10}$$

The probability of an inventory level m at retailer i for event E2

In event E2, the nominal inventory position is brought to a non-negative value by an order from retailer $j \ne i$ at time τ, $0 \le \tau < L_0$. This order is triggered by a customer demand at retailer j of the same size, and we refer to this customer demand as the *critical demand*. Lemma 3 and Corollary 2 provides the conditions under which the inventory level at retailer i is m for event E2.

Lemma 3. *For event E2, when the critical demand occurs at time* τ *(* $0 \le \tau < L_0$ *),*

$$IL_i\left(L_0 + L_i\right) = S_i - D_i\left(\tau, L_0 + L_i\right) \tag{11}$$

Corollary 2. $IL_i(L_0 + L_i) = m$ *for event E2 iff conditions a) – f) are satisfied:*

a) *a demand occurs at retailer* j *at time* τ *(the critical demand)* $\qquad , j \in \Omega \setminus [i], \ 0 \le \tau < L_0$

b) $\sum_{k \in \Omega \setminus [i,j]} D_k\left[0, \tau\right) = x$ $\qquad\qquad\qquad\qquad , x \le S_0$

c) $D_j\left[0, \tau\right) = y$ $\qquad\qquad\qquad\qquad\qquad , y \le S_0 - x$

d) $D_i\left[0, \tau\right) = z$ $\qquad\qquad\qquad\qquad\qquad , z \le S_0 - x - y$

e) *the size of the critical demand at retailer* j *is larger than* $S_0 - x - y - z$

f) $D_i\left(\tau, L_0 + L_i\right) = S_i - m$.

8

Analogously to event E1, in general conditions d) and f) in Lemma 3 are dependent. Moreover, condition c) is dependent on condition a), that there is a customer demand occurring at retailer j at τ. Again, the demands at different retailers are independent. Also, because of the independence between order sizes and inter-arrival times in the compound renewal demand process, the size of the critical demand is independent of the demand prior to τ.

The intensity at which demands occur at retailer j, in steady state, is λ_j. The probability of condition b) for a given τ and x is $\pi^\tau_{\Omega\setminus[i,j]}(x)$. When there is a demand occurrence at retailer j at τ, the probability of condition c) for a given τ and y is $\underline{\pi}^\tau_j(y)$. The joint probability of conditions d) and f) given τ and z is $\pi^{[\tau,L_0+L_i-\tau]}_i(z,S_i-m)$ and the probability of condition e) given x, y and z is $q'_j(S_0-x-y-z)$. How to determine the $\pi_{\cdot}^{\cdot}(\bullet)$-probabilities is explained in Section 3.3. Considering all possible retailers $j \neq i$, and all possible values for τ, x, y and z, the probability of an inventory level m at retailer i for event E2 at an arbitrary point in time in steady state (at time $L_0 + L_i$) can be obtained as

$$
\begin{aligned}
&P\left(IL_i\left(L_0+L_i\right)=m,E2\right) \\
&= \int_0^{L_0} \sum_{j\in\Omega\setminus[i]} \lambda_j \sum_{x=0}^{S_0} \pi^\tau_{\Omega\setminus[i,j]}(x) \sum_{y=0}^{S_0-x} \underline{\pi}^\tau_j(y) \sum_{z=0}^{S_0-x-y} q'_j(S_0-x-y-z)\pi^{[\tau,L_0+L_i-\tau]}_i(z,S_i-m)d\tau.
\end{aligned}
\tag{12}
$$

The probability of an inventory level m at retailer i for event E3

For event E3 the *critical demand* occurs at retailer i. When this demand occurs, the nominal inventory position is brought to $-\hat{B}_i$ ($\hat{B}_i > 0$). Thus, from Lemma 1 f), $\hat{B}_i$ units of the order corresponding to this critical demand (destined to retailer i) will be backordered at L_0. We refer to these $\hat{B}_i$ units as the *initial backorders*. Lemma 4 and Corollary 3 provides the conditions under which the inventory level at retailer i will be m for event E3.

Lemma 4. *For event E3, when the critical demand occurs at time τ ($0 \leq \tau < L_0$),*

$$
IL_i\left(L_0+L_i\right)=S_i-\hat{B}_i-D\left(\tau,L_0+L_i\right)
\tag{13}
$$

Corollary 3. $IL_i(L_0+L_i) = m$ *for event E3 iff conditions a) – e) are satisfied:*

a) *a demand occurs at retailer* i *at time τ (the critical demand)*   , $0 \leq \tau < L_0$

b) $\sum_{j\in\Omega\setminus[i]} D_j[0,\tau]=x$, $x \leq S_0$

c) $D_i[0,\tau]=y$, $y \leq S_0-x$

d) *the size of the critical demand is* z   , $S_0-x-y < z \leq S_i-m+S_0-x-y$

e) $D_i\left(\tau, L_0 + L_i\right) = S_i - m + S_0 - x - y - z$

Given that there is a demand (the *critical demand*) occurring at retailer i at τ, the demands in interval $[0, \tau)$ and $(\tau, L_0 + L_i)$ at retailer i are dependent on the fact that this demand occurs. However, due to the independence of the inter-arrival times and demand sizes, they are independent of the size of the *critical demand* and each other. Note again, that the demand at different retailers are independent in steady state.

The long run intensity at which demands occur at retailer i in steady state is λ_i. The probabilities for: condition b) given τ and x is $\pi^\tau_{\Omega\backslash[i]}(x)$, condition c) given τ and y is $\underline{\pi}^\tau_i(y)$, condition d) given z is $q_i(z)$ and condition e) given τ, x, y and z is $\hat{\pi}^{L_0+L_i-\tau}_i\left(S_i - m + S_0 - x - y - z\right)$. How to determine the $\pi^{\cdot}_{\cdot}(\bullet)$-probabilities is explained in Section 3.3. Considering all possible values for τ, x, y and z, the probability of an inventory level m at retailer i for event E3 at an arbitrary point in time in steady state (at time $L_0 + L_i$) can be obtained as

$$P\left(IL_i\left(L_0 + L_i\right) = m, E3\right)$$
$$= \int_0^{L_0} \lambda_i \sum_{x=0}^{S_0} \pi^\tau_{\Omega\backslash[i]}(x) \sum_{y=0}^{S_0-x} \underline{\pi}^\tau_i(y) \sum_{z=S_0-x-y+1}^{S_i-m+S_0-x-y} q_i(z)\hat{\pi}^{L_0+L_i-\tau}_i\left(S_i - m + S_0 - x - y - z\right)d\tau. \tag{14}$$

3.2. Expected Backorders Designated for Retailer i, $E[B_i(S_0)]$

Theoretically, the cost analysis and the optimization of the order-up-to levels can be performed based on only the probability mass functions of the inventory levels. However, in order to avoid evaluations of infinite sums when determining $C_i(S_0, S_i)$ $\forall i \in \Omega$ in (6), we also determine the expected amount of warehouse backorders designated for each retailer i at an arbitrary time L_0 in steady state, $E[B_i(S_0)]$. We define:

$G_i(S_0) =$ number of units ordered by retailer i in time interval $[0, L_0)$ that are dispatched from the central warehouse before time L_0.

When the expected value of $G_i(S_0)$ is known, the expected amount of backorders designated for retailer i can be determined from Lemma 5.

Lemma 5.

$$E\left[B_i\left(S_0\right)\right] = \lambda_i L_0 \mu_i - E\left[G_i\left(S_0\right)\right] \tag{15}$$

Note also that $G_i(S_0)$ only takes values between 0 and S_0, therefore

$$E\left[G_i\left(S_0\right)\right] = \sum_{g=1}^{S_0} gP\left(G_i\left(S_0\right) = g\right) \tag{16}$$

The analysis of $P(G_i(S_0) = g)$ is divided in the same three events (E1, E2 and E3) and follows a similar structure as the analysis of the inventory levels at the retailers in Section 3.1. The details of this analysis are provided in Appendix B.

3.3 Demand per time unit

This section is devoted to determining the probability mass functions of the demands during one or two consecutive time interval(s), i.e. the $\pi_\cdot^\cdot(\bullet)$-probabilities defined at the beginning of Section 3. The analysis is focused on the demand at retailer i and for notational convenience we suppress the index i from some variables. We define:

A =   inter-arrival time of customers at retailer i (with pdf $f_A(\tau)$ and cdf $F_A(\tau)$)

$V(t)$ =   time until next customer arrival after (or exactly at) time t at retailer i

$W(t)$ =   time since previous customer arrival before time t at retailer i

$U(t)$ =   inter-arrival time between the last customer arriving before time t and the first customer after (or exactly at) time t at retailer i

$q_i^n(x)$ =   probability that the sum of n customer demands at retailer i is equal to x, determined as the n-fold convolution of $q_i(x)$

$\widehat{p}_i^T(n)$ =   probability that there are n customers arriving to retailer i in time interval (t,t + T), when there is a customer arriving at t (not included in n)

$\underline{p}_i^T(n)$ =   probability that there are n customers arriving to retailer i in time interval (t,t + T), when there is a customer arriving at t + T (not included in n)

$p_i^T(n)$ =   probability that there are n customers arriving to retailer i during an arbitrary time interval of T time units in steady state

$p_i^{[T_1,T_2]}(n,m)$ = joint probability of n and m customers arriving to retailer i during two consecutive time intervals of T_1 and T_2 time units in steady state

Determining $\widehat{\pi}_i^T(x)$ and $\underline{\pi}_i^T(x)$

Note first that, due to the symmetry of the compound renewal process, $\underline{p}_i^T(x)$ is equal to $\widehat{p}_i^T(x)$, and consequently $\underline{\pi}_i^T(x) = \widehat{\pi}_i^T(x)$. Thus, we focus on determining the distribution of the demand during a time interval, when we know that there is a demand occurring at the beginning of this interval

(excluding this demand), $\widehat{\pi}_i^T(x)$. Because each customer demands at least one unit, the only possibility to have a demand of zero units is if there are no customers arriving. Thus,

$$\widehat{\pi}_i^T(0) = \widehat{p}_i^T(0) = 1 - F_A(T). \tag{17}$$

For x>0, we have to consider that there are 1 to x customers arriving in the time interval and get

$$\widehat{\pi}_i^T(x) = \sum_{n=1}^{x} \widehat{p}_i^T(n) q_i^n(x), \tag{18}$$

where the probability that there are n customers arriving can be determined recursively as

$$\widehat{p}_i^T(n) = \int_0^T f_A(\tau) \widehat{p}_i^{(T-\tau)}(n-1) d\tau. \tag{19}$$

Determining $\pi_i^T(x)$ and $\pi_\Theta^T(x)$

To determine the distribution of the demand at an arbitrary time interval from t to t + T in steady state, we study the time until the next customer arrives at retailer i after time t, V(t). See, for example, Parzen (1962) that the probability distribution function of V(t) (the excess lifetime) in steady state is

$$F_{V(t)}(v) = \frac{1}{E[A]} \int_0^v (1 - F_A(\tau)) d\tau \tag{20}$$

and the density function is

$$f_{V(t)}(v) = \frac{1 - F_A(v)}{E[A]}. \tag{21}$$

Analogously to (17), there can only be a demand of zero units in time interval [t, t + T) if no customers arrive (i.e. if the time until the next customer arrival after t is longer than T). Thus,

$$\pi_i^T(0) = p_i^T(0) = \int_T^\infty f_{V(t)}(v) dv = 1 - F_{V(t)}(T). \tag{22}$$

For x>0, we get analogously to (18) that

$$\pi_i^T(x) = \sum_{n=1}^{x} p_i^T(n) q_i^n(x). \tag{23}$$

The probability that there are n > 0 customers arriving can be obtained by conditioning on the time of the first customer arrival after time t, V(t), as

$$p_i^T(n) = \int_0^T f_{V(t)}(v) \widehat{p}_i^{(T-v)}(n-1) dv. \tag{24}$$

For determining $\pi_\Theta^T(x)$, recall that the demands at different retailers are independent in steady state. $\pi_\Theta^T(x)$ is thus determined by convolution of $\pi_i^T(x) \forall i \in \Theta$.

Determining $\pi_i^{[T_1,T_2]}(x,y)$

Finally we turn to the derivation of the joint distribution of the demands in two consecutive time intervals of lengths T_1 and T_2, $\pi_i^{[T_1,T_2]}(x,y)$. Note that whenever this probability is used in this paper, $T_1 + T_2 = L_0 + L_i$. We denote the time between the two intervals t, the time before the first interval $t_1 = t - T_1$, and the time after the second interval $t_2 = t + T_2$ (see Figure 2). We divide the analysis in four cases dependent on whether the demand is positive in each time interval: case 1: $\pi_i^{[T_1,T_2]}(0,0)$, case 2: $\pi_i^{[T_1,T_2]}(0,y)$, y>0, case 3: $\pi_i^{[T_1,T_2]}(x,0)$, x>0 and case 4: $\pi_i^{[T_1,T_2]}(x,y)$, x>0, y>0.
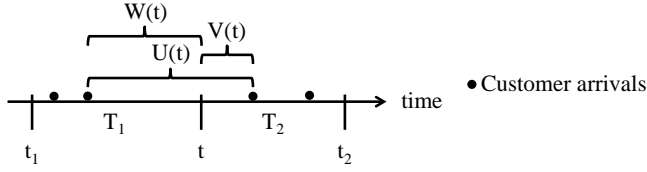

Figure 2. Timeline for analyzing case 4: $\pi_i^{[T_1,T_2]}(x,y)$, x>0, y>0

In case 1 ($\pi_i^{[T_1,T_2]}(0,0)$) the demand in both intervals is zero. As a result, analyzing the next customer arrival after t_1 analogously to (22) yields

$$\pi_i^{[T_1,T_2]}(0,0) = p_i^{[T_1,T_2]}(0,0) = p_i^{T_1+T_2}(0) = 1 - F_{V_{(t_1)}}(T_1 + T_2), \tag{25}$$

Note that the distribution of $V(t_1)$ equals the distribution of $V(t)$ (as both are in steady state) and $F_{V_{(t_1)}}(T_1 + T_2)$ can be determined by (20).

In case 2 ($\pi_i^{[T_1,T_2]}(0,y)$, $y > 0$) the demand in the first time interval (of T_1 time units) is zero and the demand in the second time interval is $y > 0$. Studying the system at time t_1, we know that the time until next customer arrival, $V(t_1)$, has to be between T_1 and $T_1 + T_2$. The probability to have zero customers arriving in the first interval and $m > 0$ customers arriving in the second interval is obtained by conditioning on $V(t_1)$

$$p_i^{[T_1,T_2]}(0,m) = \int_{T_1}^{T_1+T_2} f_{V_{(t_1)}}(v)\bar{p}_i^{T_1+T_2-v}(m-1)dv. \tag{26}$$

Analogously to (18) and (23) we get

$$\pi_i^{[T_1,T_2]}(0,y) = \sum_{m=1}^{y} p_i^{[T_1,T_2]}(0,m)q_i^m(y). \tag{27}$$

13

In case 3 ($\pi_i^{[T_1,T_2]}(x,0)$, x>0) we have a positive demand, x, in the first time interval and no demand in the second time interval. We now study the system at time t_2 and condition on the time since the previous customer arriving, $W(t_2)$, (which has to be between T_2 and $T_1 + T_2$) and get for n>0

$$p_i^{[T_1,T_2]}(n,0) = \int\limits_{T_2}^{T_1+T_2} f_{W(t_2)}(w)\underline{p}_i^{T_1+T_2-w}(n-1)dw \tag{28}$$

and

$$\pi_i^{[T_1,T_2]}(x,0) = \sum_{n=1}^{x} p_i^{[T_1,T_2]}(n,0)q_i^n(x). \tag{29}$$

Because of the symmetry of the renewal process and as both t and t_2 are in steady state, the density of $W(t_2)$ (the current lifetime) in (28), $f_{W(t_2)}(w)$, is equal to the density of V(t) in (21) (see, for instance, Karlin and Taylor, 1975, p. 193-194).

In case 4 ($\pi_i^{[T_1,T_2]}(x,y)$, x>0, y>0) we have positive demands in both time intervals. By studying the system at the time between the two time intervals, t, the analysis is based on analyzing U(t) and V(t), see Figure 2. Note first that the probability density function of the inter-arrival time between the previous and next customer arrival at retailer i at time t in steady state, U(t), (the total lifetime) is (see, for instance, Parzen, 1962)

$$f_{U(t)}(u) = \frac{uf_A(u)}{E[A]} \tag{30}$$

and the conditional density function of the time until the next customer arrival at time t, V(t), given U(t) = u is

$$f_{V(t)|U(t)=u}(v) = \begin{cases} 0 & , v > u \\ 1/u & , 0 \leq v < u \end{cases}. \tag{31}$$

Lemma 6. *At retailer* i *there are* n > 0 *customers arriving in time interval* [t_1,t) *and* m > 0 *customers arriving in time interval* [t, t_2) *iff conditions a)-d) are satisfied:*

   a) $U(t) = u$, $0 < u \leq T_1 + T_2$

   b) given a), $V(t) = v$, $\max(0, u - T_1) \leq v < \min(u, T_2)$

   c) given a) and b), m – 1 *additional customers arrive in time interval* (t + v, t_2), *excluding the customer arriving at* t + v

   d) given a) and b), n – 1 *additional customers arrive in time interval* [t_1, t + v – u), *excluding the customer arriving at* t + v – u

14

Studying the system in steady state at time t, the probability density function for condition a) is $f_{U(t)}(u)$. The conditional probability density function for condition b) is $f_{V(t)|U(t)=u}(v)$. The conditional probability for condition c) is $\widehat{p}_i^{T_2-v}(m-1)$ and for condition d) is $\underline{p}_i^{T_1+v-u}(n-1)$. Conditioning first on $U(t)=u$ and then on $V(t)=v$ and considering all possible values of u and v we get for n>0 and m>0

$$p_i^{[T_1,T_2]}(n,m) = \int_0^{T_1+T_2} f_{U(t)}(u) \int_{\max(0,u-T_1)}^{\min(u,T_2)} f_{V(t)|U(t)=u}(v)\widehat{p}_i^{T_2-v}(m-1)\underline{p}_i^{T_1+v-u}(n-1)dvdu. \tag{32}$$

From (30) and (31) (as $\max(0,u-T_1) \le v < \min(u,T_2)$) we get

$$p_i^{[T_1,T_2]}(n,m) = \frac{1}{E[A]} \int_0^{T_1+T_2} f_A(u) \int_{\max(0,u-T_1)}^{\min(u,T_2)} \widehat{p}_i^{T_2-v}(m-1)\underline{p}_i^{T_1+v-u}(n-1)dvdu. \tag{33}$$

Finally, convolution yields

$$\pi_i^{[T_1,T_2]}(x,y) = \sum_{n=1}^{x}\sum_{m=1}^{y} p_i^{[T_1,T_2]}(x,y)q_i^n(x)q_i^m(y). \tag{34}$$

4. Optimization of Order-up-to Levels

The aim of the optimization is to find the set of order-up-to levels at the warehouse, S_0^*, and all retailers, $S_1^*,\ldots, S_N^*$, that achieve the minimum total expected system costs, TC^*. It can be shown by example that the costs are not jointly convex in S_0 and S_i. However, as $C_0(S_0)$ and $C_j(S_0,S_j)$ for $j \ne i$ are independent of S_i (see the cost analysis in Section 3), the total costs are separable in the retailer order-up-to levels. Corollary 4 below also proves that the total costs are convex in S_i for a given S_0. The optimal solution is thus obtained by bounding S_0, and optimizing S_i for all $i \in \Omega$ using this convexity property. The recursive cost evaluation technique in Proposition 1 below is used to reduce the computation time. Let:

$IL_i(S_0,S_i) = $ the inventory level in steady state at retailer i, when the order-up-to level at the warehouse is S_0 and the order-up-to level at retailer i is S_i

Proposition 1. *Given* S_0, *the costs of retailer* i *can be determined recursively for increasing values of* S_i *as*

$$C_i(S_0,0) = b_i\left(E\left[B_i(S_0)\right] + \lambda_i\mu_iL_i\right), \tag{35}$$

$$C_i(S_0,S_i) = C_i(S_0,S_i-1) + (h_i+b_i)\alpha_i(S_0,S_i) - b_i, \tag{36}$$

where $\alpha_i\left(S_0,S_i\right) = P\left(IL_i\left(S_0,S_i\right) > 0\right)$ *is the ready-rate at retailer* i *and obtained as*

$$\alpha_i\left(S_0,0\right) = 0 \tag{37}$$

$$\alpha_i\left(S_0,S_i\right) = \alpha_i\left(S_0,S_i - 1\right) + P\left(IL_i\left(S_0,S_i\right) = 1\right) \tag{38}$$

Corollary 4. *given* S_0, *the inventory costs at stock point* i *and the total system costs are convex in* S_i.

By definition (in Section 2), $S_0 = 0$ is a lower bound. The upper bound is determined by Proposition 2 below. Let:

$S_0^u = $ upper bound for the optimal order-up-to level at the central warehouse

$\overline{TC} = $ lowest total cost found for all evaluated sets of order-up-to levels (updated concurrently during the optimization)

$C_i^l = $ lower bound on the expected holding and backorder costs at retailer i for any set of order-up-to levels at the warehouse and the retailer.

Proposition 2. *Defining* S_0^u *as the largest value for which*

$$C_0\left(S_o^u\right) \le \overline{TC} - \sum_{i \in \Omega} C_i^l, \tag{39}$$

no system where $S_0 > S_0^u$ *can minimize the total costs.*

Trivial C_i^l values can obviously be determined as $C_i^l = 0 \ \forall i \in \Omega$. For tighter lower bounds on the retailer costs C_i^l (leading to a tighter upper bound on the warehouse order up to level, S_0^u) see Appendix C. Note that the tightness of the upper bound can be important as the computational complexity in the cost analysis increases for higher S_0 values.

4.1. Optimization procedure

1. Determine $C_i^l \ \forall i \in \Omega$ according to the procedure in Appendix C.

2. Optimize $S_i \ \forall i \in \Omega$ for $S_0 = 0$ using the recursive procedure in Proposition 1 and the convexity property in Corollary 4. Note that $C_0(0) = 0$ and set the initial $\overline{TC}$ to the minimum total cost for $S_0 = 0$.

3. Increase S_0 one step at the time, updating the warehouse costs with (2). For each S_0, optimize S_i for all i using the recursive procedure in Proposition 1 and the convexity property in Corollary 4. Update $\overline{TC}$ as solutions with lower expected total costs are found.

4. Stop increasing S_0 when the upper bound S_0^u in Proposition 2 is reached. The optimal solution is now found and the associated lowest expected total cost is $TC^* = \overline{TC}$.

5. Numerical study

This Section presents the results of a numerical study (of two Test series) investigating the value of considering detailed inter-arrival time distributions. The system behavior is explored for either Gamma or Weibull distributed inter-arrival times (when the compounding distribution is constant or geometric). The results are compared with systems where the inter-arrival times are approximated with the commonly used exponential times. Finally, the possibility to approximate the Weibull distributions with gamma distributions (which have some computational advantages) is examined. The focus is on Test series 1, where the inter-arrival times are gamma distributed. The gamma distribution with shape parameter k and scale parameter θ has the probability density function

$$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}, \quad x > 0, k > 0, \theta > 0, \tag{40}$$

where $\Gamma(k)$ is the gamma function evaluated at k. This distribution has several interesting features. Gamma distributions with shape k=1 are exponentially distributed with $\lambda=1/\theta$ (resulting in a Poisson process) and gamma distributions where the shape parameter k is a positive integer are k-Erlang distributed. From an inventory control perspective, k-Erlang inter-arrival times occur in systems where a single customer experiences Poisson demand and orders with a continuous batch ordering policy with batch size k.

Any combination of mean, $E[A_i] > 0$, and standard deviation, $\sigma_i > 0$, of the inter-arrival times at retailer i can be represented with the gamma distribution. This should be compared with the commonly used exponential distribution, where the standard deviation always equals the mean. The resulting compound Poisson process, can therefore only model variance-to-mean ratios of the demand per time unit larger than or equal to one. With a demand process where the inter-arrival times are gamma distributed, any variance-to-mean ratio (of the demand per time unit) can be achieved.

From a computational perspective, a useful property is that the sum of n gamma distributed variables with shape k and scale θ, is gamma distributed with shape k and scale $n\theta$. Let the inter-arrival times of the customers at retailer i, A, be gamma distributed with shape k_i and scale θ_i. The time from a customer demand at retailer i until the nth next customer arrives (excluding the initial customer demand), denoted A(n), is then gamma distributed with shape k_i and scale $n\theta_i$. As a result, the numerically intensive recursive integral for n > 0 in (19) can be replaced with

$$\hat{p}_i^T(n) = F_{A(n)}(T) - F_{A(n+1)}(T). \tag{41}$$

Test series 1 consists of 60 problem scenarios. All problems involve three retailers, with $h_0 = h_i = L_i = 1 \ \forall i \in \Omega$ and $E[D_i] = [1, 1.5, 2]$, where $E[D_i] = \mu_i/E[A_i]$ is the expected demand per time unit at

retailer i. The test series includes all combinations of the following values on four variables; the lead time to the central warehouse, $L_0 = [1,2]$, the backorder costs at the retailers, $b_i = [10,100]$ $\forall i \in \Omega$, the average customer order size at the retailers, $\mu_i = [1,2,4]$, $\forall i \in \Omega$, and the coefficient of variation of the customer inter-arrival times at the retailers, $\rho_i = \sigma_i/E[A_i] = [0.2, 0.4, 0.6, 0.8, 1]$, $\forall i \in \Omega$. When $\mu_i = 1$, all order sizes are 1. When $\mu_i > 1$ a geometric compounding distribution is used, so that $q_i(x) = 1/\mu_i (1 - 1/\mu_i)^{x-1}$. For the coefficient of variation of the inter-arrival times, ρ_i, we focus on values smaller than or equal to one. Batch ordering at the customers, increased failure rate of parts and planned maintenance all lead to $\rho_i < 1$. Also, variance-to-mean ratios of the demand per time unit less than one, which cannot be modeled with compound Poisson processes, can only be modeled with $\rho_i < 1$. For gamma distributed demand, we have that $E[A_i] = k_i \theta_i$ and $\sigma_i = \theta_i \sqrt{k_i}$. Thus, we get $k_i = 1/\rho_i^2$ and $\theta_i = \rho_i^2 \mu_i / E[D_i]$.

For each problem we use the described analysis to evaluate; the optimal order-up-to levels, S_0^*, S_1^*, S_2^*, S_3^*, the corresponding minimum cost, TC^*, the proportion of inventory kept at the central warehouse in the optimal solution, Φ, the upper bound, S_0^u, and the expected relative cost increase of assuming that the inter-arrival times are exponential, ΔC^P. Let S_0^P, S_1^P, S_2^P and S_3^P be the optimal order-up to levels for the system where all inter-arrival time distributions are replaced with exponential distributions with the same mean value. Also, let TC^P be the costs of the system with gamma distributed inter-arrival times evaluated with order-up-to levels S_0^P, S_1^P, S_2^P and S_3^P. The expected relative cost increase of assuming that the inter-arrival times are exponential is then obtained as $\Delta C^P = (TC^P - TC^*)/TC^*$. The results for all problem scenarios are found in Appendix D. In Table 1, the average results for all problems with a specific value for ρ_i are presented.

Table 1. Average results for different ρ_i values in Test series 1 (gamma distributed customer inter-arrival times)

ρ_i	S_0^*	S_1^*	S_2^*	S_3^*	TC^*	Φ	S_0^u	ΔC^P
0.2	5.3	5.7	6.9	8.1	20.0	4.0 %	15.5	90.0 %
0.4	6.5	5.7	6.9	8.3	21.6	8.4 %	13.9	38.8 %
0.6	7.1	5.9	7.3	8.6	23.8	10.0 %	14.6	15.3 %
0.8	7.4	6.4	8.0	9.3	26.6	10.8 %	14.9	2.9 %
1	8.2	6.7	8.5	9.9	29.8	12.8 %	15.5	0.0 %
All	6.9	6.1	7.5	8.8	24.3	9.2%	14.9	29.4%

Table 1 shows that the average minimum total cost increases quite steadily with the coefficient of variation, from an average value of $TC^* = 20.0$ for $\rho_i = 0.2$ to an average value of $TC^* = 29.8$ for $\rho_i = 1$. Looking at the detailed results in Table D1a and D1b in Appendix D, this increase is fairly stable in all 12 cases when ρ_i is increasing from 0.2 to 1 and the other problem parameters are fixed. Next,

we turn our attention to the proportion of inventory held at the central warehouse, Φ. From previous studies of similar distribution systems under compound Poisson demand, a recurring finding is that the proportion of inventory kept at the central warehouse increases on average with the variance-to-mean ratio of the demand per time unit. It is interesting to see that the same tendency is found also when the variance in the demand is caused by variations in the inter-arrival times, see Table 1. When $\rho_i = 1$, $\Phi = 12.8\%$ on average, and when $\rho_i = 0.2$ only $\Phi = 4.0\%$ of the average inventory is kept at the central warehouse.

Focusing on the cost effects of assuming that the inter-arrival times are exponential, which is a common assumption, Table 1 shows that the relative cost increase, ΔC^P, grows rapidly when ρ_i decreases. For $\rho_i = 0.6$, the relative cost increase is already on average 15.3% and for the problems were $\rho_i = 0.2$, the corresponding solution for a system with exponential times yields costs that are on average 90.0% higher. Figure 3 presents these results separately for $\mu_i = 1$, 2 and 4. We see that, by far, the largest relative cost increases for assuming exponential inter-arrival times are found in the problems where $\mu_i = 1$ (when the demand sizes are constantly 1). For these systems, $\Delta C^P = 37.9\%$ on average already for $\rho_i = 0.6$, and for $\rho_i = 0.2$, $\Delta C^P = 246.9\%$ on average (outside the chart area in Figure 3), looking at the detailed results in Table D1a and D1b in Appendix D, we see that in all these four problem scenarios (with $\mu_i = 1$ and $\rho_i = 0.2$) the costs of the Poisson model is more than 200% higher. This can be explained by the fact that in these cases the variation in the demand is exclusively caused by the variation in the in the inter-arrival times. Also, as the variation in these systems becomes low (when ρ_i approaches zero), the costs decreases, and an overestimation of the order-up to levels by a few units has a big impact on the expected costs.
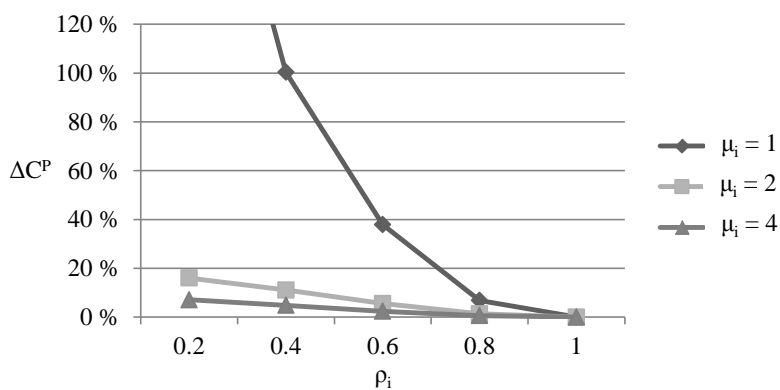

Figure 3. Relative cost increase of assuming exponential inter-arrival times, ΔC^P, as a function of ρ_i for different μ_i values.

Test series 2 explores systems where the customer inter-arrival times follow Weibull distributions. The Weibull distribution is commonly used in reliability engineering to illustrate the

failure rate of parts (see, for instance, O'Connor and Kleyner, 2012) and is therefore relevant for many spare parts systems. Note that, also the Weibull distribution is exponentially distributed when $\rho_i = 1$. The problem instances have identical parameter values (including $E[A_i]$ and ρ_i values) as Test series 1. For Test series 2, we also study the effects of approximating the Weibull distributions with gamma distributions with the same $E[A_i]$ and ρ_i values. This approximation has the computational advantage of replacing the recursive integral in (19) with (41). We define ΔC^Γ as the relative cost increase of using the gamma approximation, and it is obtained analogously to ΔC^P.

The detailed results are found in Table D2a and D2b in Appendix D. Looking at these results, it seems that it is often sufficient to approximate the Weibull distribution with a gamma distribution. The optimal policy differs from the solution with gamma distributed inter-arrival times in only 6 problem scenarios (out of 60). For five of these six problem instances, the related cost increase, ΔC^Γ, is moderate (less than 2.5%). However, in the last one, the cost increase is $\Delta C^\Gamma = 12.6\%$, proving that this approximation is not always reliable. The results from Test series 2 regarding TC^*, Φ and ΔC^P are in line with the results in Test series 1.

6. Summary and concluding remarks

We present an exact evaluation and optimization procedure for a one-warehouse-multiple-retailer inventory system with continuous base stock replenishment policies and compound renewal demand. The presented cost analysis is to the best of our knowledge the first exact analysis of the costs in a two-stage distribution system where the customer inter-arrival times at the retailers follow general continuous distributions, and we also allow for general distributions of customer demand sizes. The analysis is based on the fact that the arrival processes at the retailers, after a sufficiently long period of time, are independent of the initial state and each other. We can therefore determine the distribution of the demand in one or two consecutive time interval(s) in steady state. The inventory level distributions at the retailers can then be determined by analyzing the distribution of backorders at the central warehouse destined to a specific retailer combined with the demand during the following replenishment lead time to that retailer. A recursive cost analysis is used for optimizing the order-up-to levels in this system.

A numerical study show that it can be very costly to assume exponential customer inter-arrival times as is customary in many models, when they are, in fact, gamma or Weibull distributed. Particularly, when the coefficient of variation of the inter-arrival times $\rho_i \ll 1$ and the customer demand sizes are constantly one, this approximation performs poorly. Several problem instances where studied where the expected costs increased by more than 200% with this assumption.

Approximating Weibull inter-arrival times with gamma distributions with the same mean and standard deviation worked well for most problem instances. However, in one problem instance the relative cost increase for this assumption was as high as 12.5%.

For future research, it would be interesting to extend the analysis to other replenishment policies under compound renewal demand. Interesting replenishment policies are for instance installation stock or echelon stock (R,Q) or (s,S) ordering policies, with either continuous or periodic replenishments. It would also be interesting to see how so called *delayed ordering policies* would perform in multi-echelon settings. It has been shown in single-echelon systems, that delaying orders can create large savings when the inter-arrival times of customers are not exponential, see e.g. Axsäter and Viswanathan (2012) and Syntetos et al. (2015) and references therein. It would be interesting to see how delayed ordering could be used in multi-echelon settings.

References

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

Axsäter, S. 1993a. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* **41** 777-785.

Axsäter, S. 1993b. Optimization of order-up-to-S policies in two-echelon inventory systems with periodic review. *Naval Research Logistics* **40** 245-253.

Axsäter, S. 2000. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand. *Operations Research* **48** 686-696.

Axsäter, S., S. Viswanathan. 2012. On the value of customer information for an independent supplier in a continuous review inventory system. *European Journal of Operational Research* **221** 340-347.

Chen, F., Y.S. Zheng. 1997. One warehouse multi-retailer systems with centralized stock information. *Operations Research* **45** 275-287.

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

Forsberg, R. 1995. Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research* **81** 143-153.

Forsberg, R. 1997a. Exact evaluation of (R,Q)-policies for two-level inventory systems with Poisson demand. *European Journal of Operational Research* **96** 130-138.

Forsberg, R. 1997b. Evaluation of (R,Q)-policies for two-level inventory systems with generally distributed customer inter-arrival times. *European Journal of Operational Research* **99** 401-411.

Graves, S.C. 1996. A multi-echelon inventory model with fixed replenishment intervals. *Management Science* **42** 1-18.

Howard, C. 2013. New allocation policies for divergent inventory systems with real-time information and shipment consolidation. *Real-time Allocation Decisions in Multi-echelon Inventory Control,* Doctoral Thesis, Lund University.

Howard, C., J. Marklund. 2011. Evaluation of stock allocation policies in a divergent inventory system with shipment consolidation. *European Journal of Operational Research* **211** 298-309.

Howard, C., O. Stenius. 2014. Partial or Complete Deliveries in Two-echelon Inventory Systems? Working paper Lund University.

Karlin, K., H.M. Taylor, (1975), A First Course in Stochastic Processes, Academic Press, New York.

Kiesmüller, G.P., A.G. de Kok, S.R. Smits, P.J.M. van Laarhoven. 2004. Evaluation of divergent N-echelon (s,nQ)-policies under compound renewal demand. *OR Spectrum* **26** 547–577.

Kruse, W.K. 1981. Waiting time in a continuous review (s,S) inventory system with constant lead times. *Operations Research* **29** 202-207.

Larsen, C., A. Thorstenson. 2008. A comparison between the order and the volume fill rate for a base-stock inventory control system under a compound renewal demand process. *Journal of the Operational Research Society* **59** 798-804.

Marklund, J. 2011. Inventory Control in Divergent Supply Chains with Time-Based Dispatching and Shipment Consolidation. *Naval Research Logistics* **58** 59-71.

Marklund, J., P. Berling. 2006. Heuristic coordination of decentralized inventory systems using induced backorder costs. *Production and Operations Management* **15** 294-310.

O'Connor, P., A. Kleyner. 2012. Practical Reliability Engineering, 5th Edition. New Jersey: John Wiley & Sons.

Parzen, E. 1962. Stochastic Processes. Holden Day, San Francisco

Shang, K., Z. Tao, S. Zhou. 2015. Optimizing reorder intervals for two-echelon distribution systems with stochastic demand. Operations Research, **63** 458 – 475.

Simon, R.M. 1971. Stationary properties of a two-echelon inventory model for low demand items. *Operations Research* **19** 761-777.

Stenius, O., G. Karaarslan, J. Marklund, A.G. de Kok. 2015. Exact analysis of divergent inventory systems with time-based shipment consolidation and compound Poisson demand. Working paper Lund University and Eindhoven University of Technology.

Syntetos, A.A., R.H. Teunter, M.Z. Babai, S. Tranchel. 2015. On the benefits of delayed ordering. *European Journal of Operational Research*, forthcoming.

Appendix A – Proofs

Proof of Lemma 1:

a) Follows from (7).

b) Follows from (3) and (7) as $IL_0(L_0) = S_0 - \sum_{i \in \Omega} D_i[0, L_0] = \Psi(L_0)$.

c) The FCFS allocation ensures that before time L_0, the central warehouse can provide the first S_0 units ordered from the central warehouse in time interval $[0, L_0)$. $\Psi(\tau) \geq 0$ means that $\sum_{i \in \Omega} D_i[0, \tau] \leq S_0$, thus the statement must hold.

d) Must hold because $\Psi(\tau) \leq 0$ means that at least S_0 units have already been ordered in time interval $[0, \tau)$.

e) When $\Psi_0(\tau) = x$, $S_0 - x$ demands will have occurred in time interval $[0, \tau)$. Thus, when a demand of y units occur at time τ, x of these units will be satisfied before time L_0 and $y - x$ units of these units will be backordered at L_0.

Proof of Lemma 2: In event E1, the inventory level is non-negative at L_0. As a result, $B_i(S_0) = 0$. (9) follows from (5).

Proof of Corollary 1: Conditions a) and b) ensures and encloses all possibilities that $IL_0(L_0) \geq 0$ and c) follows from Lemma 2.

Proof of Lemma 3: From Lemma 1 c), no orders before τ will be backordered at L_0. From Lemma 1 d), $B_i(S_0) = D_i(\tau, L_0)$. (11) follows from (5).

Proof of Corollary 2: Conditions b), c) and d) ensures and encloses all possibilities that the nominal inventory position is non-negative the moment before the critical demand occurs at time τ. Conditions a) and e) ensures and encloses all possibilities that the nominal inventory position is brought to a negative value by a demand at retailer $j \neq i$ at time τ (the *critical demand*). f) follows from Lemma 3.

Proof of Lemma 4: Lemma 1 c) assures that no orders before time τ will be backordered at L_0. From Lemma 1 d) and e), it follows that $B_i(S_0) = \hat{B}_i + D_i(\tau, L_0)$. (13) then follows from (5).

Proof of Corollary 3: Conditions b) and c) ensures and encloses all possibilities that the nominal inventory position is non-negative the moment before time τ. Conditions a) and d) ensures and encloses all possibilities that the nominal inventory position is brought to a negative value by a demand at retailer i at time τ (the *critical demand*) and that $\hat{B}_i = x + y + z - S_0$ is smaller than or equal to S_i – m. e) follows from Lemma 4.

Proof of Lemma 5: No units ordered by a retailer before time 0 can be backordered at the central warehouse at time L_0. This is due to the First-Come-First-Served allocations, the constant

replenishment lead times at the warehouse, and the fact that S_0 is non-negative. Of all units ordered by retailer i in time interval $[0, L_0)$, the number of units that have been dispatched is $G_i(S_0)$, all other units are backordered at L_0. Thus $B_i(S_0) = D_i[0, L_0] - G_i(S_0)$. (15) holds as $E\left[D_i[0, L_0]\right] = \lambda_i L_0 \mu_i$.

Proof of Lemma 6: If $U(t) > T_1 + T_2$ there cannot be customers arriving in both time intervals, proving a). Per definition $V(t)$ has to be non-negative and less than $U(t) = u$. Also, $V(t)$ has to be at least $u - T_1$ in order for at least one customer to arrive in the first time interval. In order for at least one customer to arrive in the second time interval $V(t)$ has to be smaller than T_2. This proves b). c) and d) follow as the next customer after t arrives at $t + v$ and the previous customer before t arrives at $t + v - u$.

Proof of Proposition 1: (35) follows from (6). (37) holds by the definition of $\alpha_i(S_0, S_i)$. From (5) it is evident that

$$P\left(IL_i(S_0, S_i) = m\right) = P\left(IL_i(S_0, S_i + 1) = m + 1\right), \tag{A1}$$

proving (38). For any integer variable Y with a maximum value M

$$\sum_{m=1}^{M} mP(Y = m) = \sum_{n=1}^{M} P(Y \geq n). \tag{A2}$$

Thus, (36) is proven by

$$C_i(S_0, S_i) = (h_i + b_i) \sum_{m=1}^{S_i} mP\left(IL_i(S_0, S_i) = m\right) - b_i E\left[IL_i(S_0, S_i)\right]$$

$$= (h_i + b_i) \sum_{n=1}^{S_i} P\left(IL_i(S_0, S_i) \geq n\right) - b_i E\left[IL_i(S_0, S_i)\right]$$

$$= (h_i + b_i) \left(\sum_{n=2}^{S_i} P\left(IL_i(S_0, S_i) \geq n\right) + P\left(IL_i(S_0, S_i) \geq 1\right) \right) - b_i \left(E\left[IL_i(S_0, S_i - 1)\right] + 1 \right)$$

$$= (h_i + b_i) \left(\sum_{n=1}^{S_i - 1} P\left(IL_i(S_0, S_i - 1) \geq n\right) + P\left(IL_i(S_0, S_i) > 0\right) \right) - b_i E\left[IL_i(S_0, S_i - 1)\right] - b_i$$

$$= (h_i + b_i) \sum_{m=1}^{S_i - 1} mP\left(IL_i(S_0, S_i - 1) = m\right) - b_i E\left[IL_i(S_0, S_i - 1)\right] + (h_i + b_i)\alpha_i(S_0, S_i) - b_i$$

$$= C_i(S_0, S_i - 1) + (h_i + b_i)\alpha_i(S_0, S_i) - b_i,$$

where the first and last equality follows from (6), the second and fifth equality follows from (A2), and the third and fourth equality follows from (A1).

Proof of Corollary 4: From (36), $C_i(S_0, S_i) - C_i(S_0, S_i - 1) = (h_i + b_i)\alpha_i(S_0, S_i) - b_i$. This difference increases in S_i as $\alpha_i(S_0, S_i) = P\left(IL_i(S_0, S_i) > 0\right)$ increases in S_i (follows from (A1)). Also, as S_i does not affect other costs, this result holds for the total cost of the system.

Proof of Proposition 2: From (2) and (3), it follows that $C_0(S_0)$ is increasing in S_0. From the definition of S_0^u follows that $C_0(\tilde{S}_0) > \overline{TC} - \sum_{i \in \Omega} C_i^l$ for all $\tilde{S}_0 > S_0^u$. Therefore, for any $\tilde{S}_0 > S_0^u$ and any set of order up to levels, $\{S_1,...,S_N\}$,

$$TC(\tilde{S}_0,S_1,...,S_N) = C_0(\tilde{S}_0) + \sum_{i \in \Omega} C_i(\tilde{S}_0,S_i) > \overline{TC} - \sum_{i \in \Omega} C_i^l + \sum_{i \in \Omega} C_i(\tilde{S}_0,S_i) > \overline{TC} - \sum_{i \in \Omega} C_i^l + \sum_{i \in \Omega} C_i^l = \overline{TC}.$$

Thus, any system with $S_0 > S_0^u$ will have a total expected cost, TC, higher than for the best known policy, $\overline{TC}$.

Appendix B – Derivation $P(G_i(S_0) = g)$

The analysis of $P(G_i(S_0) = g)$ $(0 \leq g \leq S_0)$ is divided in the same three mutually exclusive and collectively exhaustive events defined in Section 3.1. We get

$$P(G_i(S_0) = g) = P(G_i(S_0) = g, E1) + P(G_i(S_0) = g, E2) + P(G_i(S_0) = g, E3). \tag{B1}$$

In event E1 the inventory level at time L_0 is required to be non-negative. The conditions under which $G_i(S_0) = g$ in event E1 ($IL_0(L_0)) > 0$) is specified in Lemma B1.

Lemma B1. *Of the units ordered by retailer* i *in interval* $[0,L_0)$, g *units* $(0 \leq g \leq S_0)$ *will be dispatched from the central warehouse to retailer* i *before time* L_0 *in event E1 iff conditions a) and b) are satisfied:*

a) $D_i[0,L_0) = g$

b) $\sum_{j \in \Omega \setminus [i]} D_j[0,L_0) = x$ $\qquad\qquad$, $x \leq S_0 - g$

Proof: In event E1, $IL_0(L_0) \geq 0$ which means that $B_i(S_0) = 0$ and all units ordered from the central warehouse before L_0 will be dispatched before L_0. This proofs condition a). Condition b) ensures and encloses all possibilities that $IL_0(L_0) \geq 0$. ∎

Conditions a) and b) in Lemma B1 are independent as the demands at different retailers are independent. In steady state, the probability of condition a) is $\pi_i^{L_0}(g)$ and the probability of condition b) for a given x is $\pi_{\Omega \setminus [i]}^{L_0}(x)$. How to determine these $\pi_i^{L_0}(\cdot)$-probabilities is explained in Section 3.3. Considering all possible values of x, the probability of $G_i = g$ for event E1 can be obtained as

$$P(G_i(S_0) = g, E1) = \pi_i^{L_0}(g) \sum_{x=0}^{S_0 - g} \pi_{\Omega \setminus [i]}^{L_0}(x). \tag{B2}$$

In event E2, the nominal inventory position is brought to a non-negative value by an order from retailer j ≠ i at time τ, $0 \leq \tau < L_0$ (triggered by the *critical demand*). Lemma B2 provides the conditions under which $G_i(S_0) = g$ for event E2.

Lemma B2. *Of the units ordered by retailer* i *in time interval* [0,L_0), g *units* $(0 \leq g \leq S_0)$ *will be dispatched from the central warehouse before time* L_0 *in event E2 iff conditions a)-e) are satisfied:*

a) *a demand occurs at retailer* j *at time* τ (*the critical demand*)   , $j \in \Omega \setminus [i], 0 \leq \tau < L_0$

b) $D_i[0,\tau) = g$

c) $\sum_{k \in \Omega \setminus [i,j]} D_k[0,\tau) = x$, $x \leq S_0 - g$

d) $D_j[0,\tau) = y$, $y \leq S_0 - g - x$

e) *the size of the critical demand is larger than* $S_0 - g - x - y$

Proof: Only units ordered by retailer i before the critical demand occurs at time τ will be dispatched before L_0 (see Lemma 1), which renders condition b). Conditions c) and d) ensures and encloses all possibilities that the nominal inventory position is non-negative the moment before τ. Conditions a) and e) ensures and encloses all possibilities that the nominal inventory position is brought to a negative value at time τ by an order from retailer j ≠ i. ∎

Condition d) is dependent on the fact that there is a customer demand occurring at retailer j at τ. Still, the demands at different retailers are independent, and the size of the critical demand is independent of the demand prior to τ. The intensity at which demands occur at retailer j, in the long run, is λ_j. The probability of condition b) given τ is $\pi_i^\tau(g)$ and the probability of condition c) given τ and x is $\pi_{\Omega \setminus [i,j]}^\tau(x)$. The probability of condition d) given τ and y is $\underline{\pi}_j^\tau(y)$ and the probability of condition e) given x and y is $q_j'(S_0 - g - x - y)$. How to determine the $\pi_\cdot^\cdot(\bullet)$-probabilities is explained in Section 3.3. Considering all possible retailers j ≠ i, and all possible values for τ, x and y, the probability of $G_i(S_0) = g$ for event E2 can be obtained as

$$P(G_i(S_0) = g, E2)$$
$$= \int_0^{L_0} \pi_i^\tau(g) \sum_{j \in \Omega \setminus [i]} \lambda_j \sum_{x=0}^{S_0 - g} \pi_{\Omega \setminus [i,j]}^\tau(x) \sum_{y=0}^{S_0 - g - x} \underline{\pi}_j^\tau(y) q_j'(S_0 - x - y - z) d\tau. \tag{B3}$$

In event E3 the *critical demand* occurs at retailer i. Note that if the nominal inventory position is brought from a positive value x to a negative value, x units of this order will be dispatched from the central warehouse before time L_0. Lemma B3 provides the conditions under which $G_i(S_0) = g$ for event E3.

Lemma B3. *Of all units ordered by retailer* i *in time interval* [0,L_0), g *units* ($0 \leq g \leq S_0$) *will be dispatched from the central warehouse before time* L_0 *in event E3 iff conditions a)-d) are satisfied:*

a) *a demand occurs at retailer i at time* τ (*the critical demand*)   , $0 \leq \tau < L_0$

b) $\sum_{j \in \Omega \backslash [i]} D_j[0,\tau] = S_0 - g$

c) $D_i[0,\tau] = y$, $0 \leq y \leq g$

d) *the size of the critical demand is larger than* g – y

Proof: As the inventory level at the central warehouse is negative at L_0 in Event E3, the amount of units demanded in time interval [0,L_0) delivered to all retailers is S_0. Lemma 1 gives that the units delivered to all retailers except retailer i are the ones demanded before time τ, proving condition b). Condition c) ensures and encloses all possibilities that the nominal inventory position is non-negative the moment before τ. Conditions a) and d) ensures and encloses all possibilities that the nominal inventory position is brought to a negative value at time τ by an order from retailer i. ∎

Again, the demand in time interval [0,τ), prior to the critical demand is dependent on the fact that the critical demand occurs, but it is independent of the size of this demand at time τ. The long run intensity at which demands occur at retailer i is λ_i. The probabilities for condition b) given τ is $\pi^\tau_{\Omega \backslash [i]}(S_0 - g)$, condition c) given τ and y is $\underline{\pi}^\tau_i(y)$, and condition d) given y is $q'_i(g - y)$. How to determine the $\pi^\cdot_\cdot(\bullet)$-probabilities is explained in Section 3.3. Considering all possible values for τ and y, the probability for $G_i(S_0) = g$ for event E3 can be obtained as

$$P\big(G_i(S_0) = g, E3\big)$$
$$= \int_0^{L_0} \lambda_i \pi^\tau_{\Omega \backslash [i]}(S_0 - g) \sum_{y=0}^{g} \underline{\pi}^\tau_i(y) q'_i(g - y) d\tau. \tag{B4}$$

Appendix C – Lower bound for expected costs at retailer i

For systems with compound Poisson demand at retailer i, a lower bound for the expected costs at retailer i, C^1_i, can be determined as the lowest cost for the single echelon system of retailer i, where the central warehouse can fulfill all retailer orders immediately (see Stenius et al., 2015). This is not true for all compound renewal systems. To see this, consider, for example, a system consisting of a warehouse and a single retailer. The retailer faces renewal demand (customer demand sizes are constantly 1) where the inter-arrival times have a mean 2 and a standard deviation approaching zero. The other parameter values are: $L_0 = L_1 = h_0 = h_1 = 1$ and $b_1 = 10$. The minimum expected total system costs for the two-echelon system approaches zero and can be achieved when S_0 is set to 0 and

S_1 is set to 1. However, for the single-echelon system, the minimum costs for the single echelon system is achieved with $S_1 = 1$ and is approximately 0.5. Thus, in this special case, the order delay caused by the central warehouse replenishment lead time actually decreases the total costs of the system.

For determining a lower bound applicable for any compound renewal demand structure we again study the system at an arbitrary time t in steady state, and for notational convenience we let t = $L_0 + L_i$. The influence of the warehouse on the retailer performance can be summarized by the distribution of the backorders at the central warehouse at time L_0 destined to retailer i, $B_i(S_0)$, keeping in mind that it can be dependent on the time since the previous customer arrival at retailer i at time $L_0, W(L_0)$. The lower bound is thus achieved by conditioning on $B_i(S_0)$ and $W(L_0)$ and artificially optimizing S_i at time L_0 (i.e. S_i is allowed to depend on x and w) in Lemma C1 below. Let

$C_i\left(S_i \mid B_i\left(S_0\right) = x, W\left(L_0\right) = w\right)$ = Expected cost per time unit at retailer i at time $L_0 + L_i$ for reorder
    level S_i, given $B_i(S_0) = x$ and $W(L_0) = w$

$IL_i\left(S_i \mid B_i\left(S_0\right) = x, W\left(L_0\right) = w\right)$ = Inventory level at retailer i at time $L_0 + L_i$ for reorder level S_i,
    given $B_i(S_0) = x$ and $W(L_0) = w$

$\pi_i^{L_i \mid W(L_0)=w}\left(x\right)$ = probability of x units demanded at retailer i in time interval $[L_0, L_0 + L_i)$, given
    $W(L_0) = w$

$p_i^{L_i \mid W(L_0)=w}\left(n\right)$ = probability of n customers arriving to retailer i in time interval $[L_0, L_0 + L_i)$, given
    $W(L_0) = w$

Lemma C1. *For any* S_0 *and* S_i

$$C_i\left(S_0, S_i\right) \geq \int_0^\infty f_{W(L_0)}\left(w\right) \min_{s_i} C_i\left(s_i \mid B_i\left(S_0\right) = 0, W\left(L_0\right) = w\right) dw. \qquad (C1)$$

Proof: From (5) follows that $IL_i\left(S_i \mid B_i\left(S_0\right) = x, W\left(L_0\right) = w\right) = IL_i\left(S_i - x \mid B_i\left(S_0\right) = 0, W\left(L_0\right) = w\right)$.

Thus, $\qquad C_i\left(S_i \mid B_i\left(S_0\right) = x, W = w\right) = C_i\left(S_i - x \mid B_i\left(S_0\right) = 0, W = w\right)$, assuring that

$\min_{s_i} C_i\left(s_i \mid B_i\left(S_0\right) = x, W = w\right) = \min_{s_i} C_i\left(s_i \mid B_i\left(S_0\right) = 0, W = w\right)$. Let $f_{B_i(S_0), W(L_0)}\left(x, w\right)$ be the joint

probability density function of $B_i(S_0)$ and $W(L_0)$. Conditioning on both $B_i(S_0)$ and $W(L_0)$ gives

$$C_i(S_0, S_i) = \sum_{x=0}^{\infty} \int_0^{\infty} f_{B_i(S_0), W(L_0)}(x, w) C_i\left(S_i \big| B_i(S_0) = x, W(L_0) = w\right) dw$$

$$\geq \sum_{x=0}^{\infty} \int_0^{\infty} f_{B_i(S_0), W(L_0)}(x, w) \min_{s_i} C_i\left(s_i \big| B_i(S_0) = x, W(L_0) = w\right) dw \qquad (C2)$$

$$= \int_0^{\infty} f_{W(L_0)}(w) \min_{s_i} C_i\left(s_i \big| B_i(S_0) = 0, W(L_0) = w\right) dw.$$

∎

In order to avoid the infinite integral in (C1), we introduce a large value, M_1, and stop evaluating the integral at this value (i.e. costs are set to 0 for higher values). Thus, this new integral will have a smaller value than the original. Also, evaluating

$$C_i\left(s_i \big| B_i(S_0) = 0, W(L_0) = w\right) dw =$$
$$h_i E\left[IL_i^+\left(s_i \big| B_i(S_0) = 0, W(L_0) = w\right)\right] + b_i E\left[IL_i^-\left(s_i \big| B_i(S_0) = 0, W(L_0) = w\right)\right] \qquad (C3)$$

is numerically challenging as it either includes an infinite sum or requires the determination of $E\left[IL\left(s_i \mid B_i(S_0) = 0, W(L_0) = w\right)\right]$. Finding a generic expression for this expected value is difficult, we therefore define $\widetilde{C}_i\left(S_i \big| B_i(S_0) = 0, W(L_0) = w, M_2, M_3\right)$ to be $C_i\left(S_i \big| B_i(S_0) = 0, W(L_0) = w\right)$ with the distinction that all situations where more than M_2 customers arrives in $[L_0, L_0 + L_i)$ and all situations where an individual customer orders more than M_3 units are disregarded (costs are set to 0 in these cases). It is easy to show that $C_i\left(S_i \big| B_i(S_0) = 0, W(L_0) = w\right) \geq \widetilde{C}_i\left(S_i \big| B_i(S_0) = 0, W(L_0) = w, M_2, M_3\right) \forall M_2, M_3$. These properties, together with Lemma C1, prove Proposition C1.

Proposition C1. *The expected retailer costs at retailer* i *for any values of* S_0 *and* S_i *and any* M_1, M_2 *and* M_3, *is always at least*

$$C_i^l = \int_0^{M_1} f_{W(L_0)}(w) \min_{s_i} \widetilde{C}_i\left(s_i \big| B_i(S_0) = 0, W(L_0) = w, M_2, M_3\right) dw. \qquad (C4)$$

For retailers where the inter-arrival rates are exponential this lower bound on the retailer costs approaches the actual minimum retailer costs when M_1, M_2 and M_3 approach infinity. When deciding on M_1, M_2 and M_3 one should keep in mind that higher values will create a tighter bound on the retailer costs but will be computationally slower. A suggestion for deciding these values is to first decide on γ_1, γ_2 and γ_3 values (for instance $\gamma_1 = \gamma_2 = \gamma_3 = 99\%$). M_1, M_2 and M_3 is then determined as

$$F_{W(L_0)}(M_1) = \gamma_1, \qquad (C5)$$

$$\min\left\{ M_2 : \sum_{n=0}^{M_2 - 1} \widehat{p}_i^{L_i}(n) \geq \gamma_2 \right\} \quad \text{and} \qquad (C6)$$

29

$$\min\left\{M_3 : \sum_{x=0}^{M_3} q_i(x) \geq \gamma_3\right\}. \tag{C7}$$

Thus, e.g. M_1 is determined so that γ_1 of the probability mass of the integral in (C1) is included in (C4). Note that, $W(L_0)$ has the same unconditioned distribution as $V(t)$ and (C5) can be determined by (20).

For deciding $\tilde{C}_i\left(s_i \middle| B_i(S_0) = 0, W = w, M_2, M_3\right)$ we need to determine $P\left(IL_i(S_i) = m \middle| B_i(S_0) = 0, W(L_0) = w\right)$. With $B_i(S_0) = 0$, this probability follows from (5)

$$P\left(IL_i(S_i) = m \middle| B_i(S_0) = 0, W(L_0) = w\right) = \pi_i^{L_i|W(L_0)=w}(S_i - m). \tag{C8}$$

Analogous to the previous π-functions, we get for $S_i - m > 0$,

$$\pi_i^{L_i|W(L_0)=w}(S_i - m) = \sum_{n=1}^{S_i - m} p_i^{L_i|W(L_0)=w}(n) q_i^n(S_i - m), \tag{C9}$$

where

$$p_i^{L_i|W(L_0)=w}(n) = \int_0^{L_i} f_{V(L_0)|W(L_0)=w}(v) \hat{p}_i^{(L_i - v)}(n - 1) dv. \tag{C10}$$

Finally, $f_{V(L_0)|W(L_0)=w}(v)$ is obtained from Lemma C2 and the case where $S_i - m = 0$, i.e. $\pi_i^{L_i|W(L_0)=w}(0) = p_i^{L_i|W(L_0)=w}(0)$ is determined by Corollary C1.

Lemma C2.

$$f_{V(L_0)|W(L_0)=w}(v) = \frac{f_A(v + w)}{1 - F_A(w)}. \tag{C11}$$

Proof: $W(L_0)$ has the same unconditioned density function as $V(t)$ in (21). Also, analogously to (31) $f_{W(L_0)|U(L_0)=v+w}(w) = 1/(v + w)$ for positive values on v and w. These facts together with theory on conditional probabilities prove that

30

$$f_{V(L_0)|W(L_0)=w}(v)$$

$$= f_{U(L_0)|W(L_0)=w}(v+w)$$

$$= \frac{f_{U(L_0),W(L_0)}(v+w,w)}{f_{W(L_0)}(w)}$$

$$= \frac{f_{W(L_0)|U(L_0)=v+w}(w)f_{U(L_0)}(v+w)}{f_{W(L_0)}(w)} \tag{C12}$$

$$= \frac{\dfrac{1}{(v+w)}\dfrac{(v+w)f_A(v+w)}{E[A]}}{\dfrac{1-F_A(w)}{E[A]}}$$

$$= \frac{f_A(v+w)}{1-F_A(w)}.$$

■

Corollary C1.

$$\pi_i^{L_i|W(L_0)=w}(0) = p_i^{L_i|W(L_0)=w}(0) = \frac{1-F_A(L_i+w)}{1-F_A(w)}. \tag{C13}$$

Proof: No demand in time interval $[L_0, L_0 + L_i)$ implies that no customers arrive, thus from Lemma C2 follows

$$\pi_i^{L_i|W(L_0)=w}(0) = p_i^{L_i|W(L_0)=w}(0)$$

$$= \int_{L_i}^{\infty} f_{V(L_0)|W(L_0)=w}(v)dv = \int_{L_i}^{\infty} \frac{f_A(v+w)}{1-F_A(w)}dv = \frac{1-F_A(v+w)}{1-F_A(w)} \tag{C14}$$

■

Appendix D – Results of the numerical study

Table D1a. Results of Test series 1, with gamma distributed customer inter-arrival times, presented in Section5. First 45 problem scenarios.

#	L_0	b_i	μ_i	ρ_i	S_0^*	S_1^*	S_2^*	S_3^*	TC^*	Φ	S_0^u	ΔC^P
1	1	10	1	0.2	2	2	3	4	2.3	0.0 %	6	201.6 %
2	1	10	1	0.4	3	2	3	4	3.7	0.9 %	5	87.8 %
3	1	10	1	0.6	4	2	3	4	5.4	7.9 %	5	32.5 %
4	1	10	1	0.8	5	2	3	4	7.4	18.8 %	6	5.8 %
5	1	10	1	1	4	3	4	5	9.2	8.2 %	6	-
6	1	10	2	0.2	4	3	4	5	12.9	9.8 %	9	10.2 %
7	1	10	2	0.4	3	4	5	6	13.8	4.2 %	8	6.8 %
8	1	10	2	0.6	4	4	5	6	15.2	8.8 %	9	3.3 %
9	1	10	2	0.8	4	4	5	6	17.0	9.7 %	9	0.6 %
10	1	10	2	1	4	4	6	7	19.1	9.1 %	9	-
11	1	10	4	0.2	1	6	8	9	25.3	1.5 %	16	4.7 %
12	1	10	4	0.4	3	6	7	9	26.4	6.1 %	13	2.6 %
13	1	10	4	0.6	4	6	7	9	28.0	9.0 %	13	1.1 %
14	1	10	4	0.8	4	6	8	9	30.0	8.9 %	13	0.3 %
15	1	10	4	1	5	6	8	9	32.4	12.1 %	13	-
16	1	100	1	0.2	3	2	3	4	3.1	0.2 %	6	284.4 %
17	1	100	1	0.4	5	2	3	4	5.6	14.1 %	7	113.1 %
18	1	100	1	0.6	4	3	4	5	8.4	4.7 %	8	43.9 %
19	1	100	1	0.8	4	4	5	6	11.5	4.6 %	8	7.9 %
20	1	100	1	1	6	4	5	6	14.8	15.2 %	8	-
21	1	100	2	0.2	4	7	8	9	23.5	3.9 %	14	19.3 %
22	1	100	2	0.4	5	7	8	9	24.9	6.9 %	13	13.5 %
23	1	100	2	0.6	5	7	9	10	27.2	6.8 %	13	7.0 %
24	1	100	2	0.8	5	8	10	11	30.2	6.4 %	13	1.7 %
25	1	100	2	1	6	8	10	12	33.8	8.9 %	14	-
26	1	100	4	0.2	2	14	15	17	49.0	1.5 %	26	8.0 %
27	1	100	4	0.4	4	13	15	17	50.7	3.9 %	20	5.4 %
28	1	100	4	0.6	6	13	15	17	53.4	7.0 %	21	2.7 %
29	1	100	4	0.8	6	14	16	18	57.2	6.7 %	21	0.6 %
30	1	100	4	1	7	14	17	19	62.0	8.0 %	22	-
31	2	10	1	0.2	6	2	3	4	2.3	0.0 %	10	226.7 %
32	2	10	1	0.4	8	2	3	4	4.1	4.7 %	10	81.7 %
33	2	10	1	0.6	9	2	3	4	6.0	16.1 %	10	28.6 %
34	2	10	1	0.8	8	3	4	5	8.2	8.0 %	11	3.9 %
35	2	10	1	1	9	3	4	5	10.2	15.3 %	11	-
36	2	10	2	0.2	8	4	5	6	13.8	8.3 %	14	12.6 %
37	2	10	2	0.4	8	4	5	6	14.9	9.5 %	14	8.4 %
38	2	10	2	0.6	9	4	5	6	16.5	14.7 %	14	3.9 %
39	2	10	2	0.8	9	4	6	7	18.6	14.0 %	15	0.8 %
40	2	10	2	1	10	4	6	7	21.0	18.6 %	16	-
41	2	10	4	0.2	7	6	8	9	27.4	7.5 %	22	5.0 %
42	2	10	4	0.4	8	6	8	9	28.5	10.4 %	19	3.2 %
43	2	10	4	0.6	9	6	8	9	30.4	13.6 %	20	1.4 %
44	2	10	4	0.8	10	6	8	10	32.8	16.1 %	20	0.3 %
45	2	10	4	1	11	6	8	10	35.7	19.1 %	21	-

Table D1b. Results of Test series 1, with gamma distributed customer inter-arrival times, presented in Section5. Last 15 problem scenarios.

#	L_0	b_i	μ_i	ρ_i	S_0^*	S_1^*	S_2^*	S_3^*	TC^*	Φ	S_0^u	ΔC^P
46	2	100	1	0.2	7	2	3	4	3.6	0.1 %	11	274.9 %
47	2	100	1	0.4	10	2	3	4	6.2	21.4 %	12	118.8 %
48	2	100	1	0.6	9	3	4	5	9.2	9.9 %	13	46.5 %
49	2	100	1	0.8	10	4	5	6	12.7	13.6 %	14	9.8 %
50	2	100	1	1	10	4	6	7	16.3	13.1 %	14	-
51	2	100	2	0.2	10	7	8	9	24.8	9.4 %	19	22.0 %
52	2	100	2	0.4	10	7	8	10	26.4	9.6 %	19	15.7 %
53	2	100	2	0.6	11	7	9	10	29.0	12.4 %	20	8.0 %
54	2	100	2	0.8	11	8	10	11	32.4	11.7 %	20	2.1 %
55	2	100	2	1	12	9	10	12	36.5	13.6 %	21	-
56	2	100	4	0.2	9	13	15	17	51.8	5.8 %	33	10.7 %
57	2	100	4	0.4	11	13	15	17	53.5	8.8 %	27	8.0 %
58	2	100	4	0.6	11	14	16	18	56.7	8.6 %	29	4.4 %
59	2	100	4	0.8	13	14	16	19	61.0	11.4 %	29	1.4 %
60	2	100	4	1	14	15	18	20	66.4	12.2 %	31	-

When deciding C_i^1, $\gamma_1 = \gamma_2 = \gamma_3 = 99\%$. Integrations are performed with the trapezoidal method with integration interval 0.001 for tabulating values and initial calculations and integration interval 0.01 for the computationally heavy parts. The analytical method has been validated through discrete event simulations.

Table D2a. Results of Test series 2, with Weibull distributed customer inter-arrival times, presented in Section5. First 25 problem scenarios.

#	L_0	b_i	μ_i	ρ_i	S_0^*	S_1^*	S_2^*	S_3^*	TC^*	Φ	S_0^u	ΔC^P	ΔC^I
1	1	10	1	0.2	2	2	3	4	2.4	0.0%	6	188.8 %	-
2	1	10	1	0.4	3	2	3	4	4.0	0.9%	5	77.1 %	-
3	1	10	1	0.6	4	2	3	4	5.7	8.1%	6	27.6 %	-
4	1	10	1	0.8	5	2	3	4	7.6	18.9%	6	4.3 %	-
5	1	10	1	1	4	3	4	5	9.2	8.2%	6	-	-
6	1	10	2	0.2	4	3	4	5	12.9	9.9%	9	10.0 %	-
7	1	10	2	0.4	3	4	5	6	13.9	4.3%	8	6.5 %	-
8	1	10	2	0.6	4	4	5	6	15.4	8.9%	8	2.9 %	-
9	1	10	2	0.8	4	4	5	7	17.2	9.0%	9	0.4 %	0.02 %
10	1	10	2	1	4	4	6	7	19.1	9.1%	9	-	-
11	1	10	4	0.2	1	6	8	9	25.4	1.5%	15	4.6 %	-
12	1	10	4	0.4	3	6	7	9	26.7	6.2%	13	2.4 %	-
13	1	10	4	0.6	4	6	7	9	28.4	9.1%	12	0.9 %	-
14	1	10	4	0.8	4	6	8	9	30.3	9.0%	13	0.2 %	-
15	1	10	4	1	5	6	8	9	32.4	12.1%	13	-	-
16	1	100	1	0.2	3	2	3	4	3.4	0.2%	7	252.4 %	-
17	1	100	1	0.4	5	2	3	4	6.4	14.3%	7	88.7 %	-
18	1	100	1	0.6	5	3	4	5	9.0	10.7%	8	34.2 %	1.3 %
19	1	100	1	0.8	4	4	5	6	11.9	4.7%	8	5.8 %	-
20	1	100	1	1	6	4	5	6	14.8	15.2%	8	-	-
21	1	100	2	0.2	4	7	8	9	23.5	3.9%	13	19.0 %	-
22	1	100	2	0.4	5	7	8	9	25.3	7.0%	13	12.4 %	-
23	1	100	2	0.6	5	7	9	10	27.7	6.9%	13	5.8 %	-
24	1	100	2	0.8	5	8	10	11	30.6	6.5%	13	1.4 %	-
25	1	100	2	1	6	8	10	12	33.8	8.9%	14	-	-

Table D2b. Results of Test series 2, with Weibull distributed customer inter-arrival times, presented in Section 5. Last 35 problem scenarios and average over all problem scenarios.

#	L_0	b_i	μ_i	ρ_i	S_0^*	S_1^*	S_2^*	S_3^*	TC^*	Φ	S_0^u	ΔC^P	ΔC^Γ
26	1	100	4	0.2	2	14	15	17	49.1	1.5%	24	7.8 %	-
27	1	100	4	0.4	4	13	15	17	51.3	4.0%	20	4.9 %	-
28	1	100	4	0.6	6	13	15	17	54.3	7.0%	21	2.1 %	-
29	1	100	4	0.8	6	14	16	18	58.0	6.7%	22	0.4 %	-
30	1	100	4	1	7	14	17	19	62.0	8.0%	22	-	-
31	2	10	1	0.2	6	2	3	4	2.4	0.0%	10	208.3 %	-
32	2	10	1	0.4	8	2	3	4	4.3	4.6%	10	74.0 %	-
33	2	10	1	0.6	9	2	3	4	6.2	16.2%	11	24.6 %	-
34	2	10	1	0.8	8	3	4	5	8.4	8.0%	11	3.5 %	-
35	2	10	1	1	9	3	4	5	10.2	15.3%	11	-	-
36	2	10	2	0.2	8	4	5	6	13.8	8.4%	14	12.6 %	-
37	2	10	2	0.4	8	4	5	6	15.0	9.6%	14	8.1 %	-
38	2	10	2	0.6	9	4	5	6	16.8	14.8%	14	3.5 %	-
39	2	10	2	0.8	9	4	6	7	18.9	14.1%	15	0.7 %	-
40	2	10	2	1	10	4	6	7	21.0	18.6%	16	-	-
41	2	10	4	0.2	7	6	8	9	27.5	7.6%	22	4.9 %	-
42	2	10	4	0.4	8	6	8	9	28.8	10.5%	19	3.0 %	-
43	2	10	4	0.6	9	6	8	9	30.9	13.7%	19	1.2 %	-
44	2	10	4	0.8	10	6	8	10	33.2	16.2%	20	0.2 %	-
45	2	10	4	1	11	6	8	10	35.7	19.1%	21	-	-
46	2	100	1	0.2	8	2	3	4	3.8	1.3%	12	255.7 %	12.6 %
47	2	100	1	0.4	10	2	3	4	6.9	21.5%	12	95.7 %	-
48	2	100	1	0.6	10	3	4	5	9.8	16.1%	14	39.0 %	2.5 %
49	2	100	1	0.8	10	4	5	6	13.0	13.7%	14	8.2 %	-
50	2	100	1	1	10	4	6	7	16.3	13.1%	14	-	-
51	2	100	2	0.2	10	7	8	9	24.8	9.5%	19	21.8 %	-
52	2	100	2	0.4	10	7	8	10	26.7	9.6%	19	14.6 %	-
53	2	100	2	0.6	10	8	9	11	29.6	9.1%	20	6.8 %	-
54	2	100	2	0.8	11	8	10	11	32.8	11.7%	20	1.6 %	-
55	2	100	2	1	12	9	10	12	36.5	13.6%	21	-	-
56	2	100	4	0.2	9	13	15	17	51.8	5.8%	31	10.6 %	-
57	2	100	4	0.4	11	13	15	17	54.1	8.9%	27	7.4 %	-
58	2	100	4	0.6	12	14	16	18	57.6	10.0%	28	3.7 %	0.02 %
59	2	100	4	0.8	13	14	17	19	61.8	11.3%	30	1.1 %	0.06 %
60	2	100	4	1	14	15	18	20	66.4	12.2%	31	-	-
Average	1.5	55	2.3	0.6	6.9	6.1	7.6	8.9	24.6	9.4 %	14.8	26.2 %	0.3 %

Knowing $E[A_i]$ and ρ_i the corresponding Weibull distributions are retrieved with Microsoft Excel's solver function. When deciding C_i^1, $\gamma_1 = \gamma_2 = \gamma_3 = 99\%$. Integrations are performed with the trapezoidal method with integration interval 0.0001 for tabulating values and initial calculations (including the recursive integral in (19)) and integration interval 0.01 for the computationally heavy parts. The analytical method has been validated through discrete event simulations.

Paper V

A Note on Solution Procedures for a Class of Two-echelon Inventory Problems

Olof Stenius

Department of Industrial Management and Logistics, Lund University
olle.stenius@iml.lth.se

Abstract

The literature contains several exact models for inventory control of stochastic multi-echelon systems that determine the expected costs directly, without first deriving the distribution of the inventory levels, which is the standard approach. This technical note extends the scope of these models by showing how the inventory level distributions can be obtained from the cost functions for a broad class of publications. This facilitates the analysis of performance measures, including the ready rate and the fill rate for these systems. We also show that a well-known relationship between the ready rate and the cost parameters is valid for many stock points in multi-echelon systems.

1. Introduction and Related Literature

The multi-echelon inventory literature contains a wide range of models for exact cost evaluation of divergent systems, which are not based on first determining the inventory level distributions. These models, originating with Axsäter (1990), are based on following a unit through the system and determining the costs this unit accrues at different stages. In this note we show how the inventory level distributions for a wide range of policies may easily be determined from the expected costs. This offers a way to extend the scope of analysis for many of these existing models to performance measures based on inventory level information. The inventory level distribution is, for instance, used for computing the ready rate (the proportion of time with positive stock on hand) and the fill rate (the proportion of demand that can be satisfied immediately from stock on hand, see, for example, Axsäter 2006, p.97-98). Moreover, this note shows that a well-known relationship between the ready rate and the holding and backorder costs is valid for many stock points in multi-echelon systems. It also provides an alternative convexity proof of the costs at the retailers in the retailer order-up-to levels. For expositional reasons, the results are initially explained in the context of Axsäter (1990) and then generalized.

The inventory system considered in Axsäter (1990) consists of a central warehouse that supplies a number of retailers facing Poisson demand. All stock points are subject to linear holding costs, and linear backorder costs are incurred at all retailers. There are continuous (S-1,S) replenishment policies (one-for-one replenishments), complete backordering, FCFS allocations and deterministic replenishment times at all locations (see Axsäter, 1990, for motivations and a detailed problem formulation).

The presented results are generalized to a group of models that offer exact cost analysis of other types of multi-echelon inventory systems; Axsäter (1993a,b, 1997, 1998), Forsberg (1995, 1997a,b) and Marklund (2002, 2011). More specifically, Forsberg (1995) extends the analysis of Axsäter (1990) to compound Poisson demand (where each customer orders a random number of units). For batch ordering policies, Axsäter (1993b) extends the Poisson analysis to the case of installation stock (R,Q) policies at all stock points with identical retailers. This is generalized to non-identical retailers by Forsberg (1997a) and Axsäter (1998). Forsberg (1997b) generalizes this model to the case where the time between two consecutive customers is Erlang distributed. Axsäter (1997) considers the case of echelon stock (R,Q) policies (the warehouse also considers the stock levels at the retailers when placing orders) in combination with compound Poisson demand. Marklund (2002) presents an alternative way to consider the information at the retailers. He introduces an (α_0, Q_0) replenishment

policy at the central warehouse, which aims at synchronizing the incoming replenishments with the expected time of future retailer orders.

Axsäter (1993a) provides the analysis for periodic order-up-to-S policies. Forsberg (1995) analyzes a periodic replenishment system with compound Poisson demand. Marklund (2011) distinguishes the ordering policy from the replenishment policy and combines a continuous (R,Q) ordering policy at the central warehouse and (S-1,S) ordering policies at the retailers (facing Poisson demand) with consolidated periodic replenishments to groups of retailers from the warehouse.

None of the models above evaluate the inventory level distributions at the stock points. There are however several models that do. We focus here on models analyzing the same systems as the cited papers above. Simon (1971) presents an exact analysis of the inventory level distributions for the same system as Axsäter (1990). For echelon-stock (R,Q) policies under Poisson demand, these distributions are evaluated in Chen and Zheng (1997). In a parallel work, Axsäter (2000) presents a different evaluation technique for installation stock (R,Q) policies and compound Poisson demand. Graves (1996) determines the inventory level distributions for a system with periodic replenishments and order-up-to S policies. The system analyzed in Marklund (2011) is extended to compound Poisson demand and evaluation of inventory levels in Stenius et al. (2016). Stenius (2016) analyzes a continuous review (S-1,S) replenishments system where the customer demand is compound renewal (i.e. the inter-arrival times can follow any continuous distribution and the demand sizes any discrete distribution).

As can be seen, the inventory level distribution is analyzed for many of the models for which the main result of this note (the derivation of the inventory level from the cost function) can be applied. This note provides the inventory level distribution for the remaining models in Forsberg (1995, 1997b), Axsäter (1997) and Marklund (2002). Furthermore, it presents an alternative way of analysis for the inventory levels in the other models. It can also benefit future models by providing the inventory level distributions from the cost functions directly. The presented results regarding the relationship between the ready-rate and the holding and backorder costs, as well as the convexity proof is valid for a large group of stock points, e.g. models in all papers cited above.

This note is organized as follows; Section 2 specifies the general system requirements and explains why the Axsäter (1990) model satisfies these requirements. The analysis of the relationships between the inventory level distributions and the costs for systems fulfilling these requirements is presented in Section 3. Section 4 generalizes the results to other problem settings and Section 5 presents some concluding remarks.

2. System Specifications

This section presents the general requirements for the considered stock point and the inventory system it is part of, for which the results in Section 3 are valid. We also explain why these requirements are satisfied by the retailers in the Axsäter (1990) model.

Consider a stock point i with an inventory level that only takes integer values (with a finite maximum value). At this stock point, there are linear holding costs, $h_i > 0$, and backorder costs, $\beta_i \geq 0$, per unit and time unit. Let

S_i = *adjusting control variable* for stock point i, defined below

$\mathbf{V}_i$ = vector of all other control variables affecting the long run inventory level distribution at stock point i

$IL_i(\mathbf{V}_i,S_i)$ = inventory level at stock point i (= stock on hand – backorders)

x^+ = max(x,0), $x^- = $ max(–x,0)

The *adjusting control variable* S_i is adjusting in the sense that an alteration of S_i with Δ units, adjusts the long run inventory level distribution of stock point i with Δ units. More specifically, for the *adjusting control variable* for stock point i, S_i, the following three conditions must hold:

1. There is a cost function related to the inventory level distribution at stock point i of the form

$$C_i\left(\mathbf{V}_i,S_i\right) \ = \ h_i E\left[IL_i^+\left(\mathbf{V}_i,S_i\right)\right] + \beta_i E\left[IL_i^-\left(\mathbf{V}_i,S_i\right)\right] \tag{1}$$

for $h_i > 0$ and $\beta_i \geq 0$.

2. A shift in S_i with Δ units shifts the long run inventory level distribution of stock point i with the same amount, i.e.

$$P\left(IL_i\left(\mathbf{V}_i,S_i\right) = m\right) = P\left(IL_i\left(\mathbf{V}_i,S_i + \Delta\right) = m + \Delta\right), \quad \forall m \tag{2}$$

3. S_i does not affect any other expected system costs than the inventory costs at stock point i, $C_i\left(\mathbf{V}_i,S_i\right)$.

For the model in Axsäter (1990), the order-up-to level, S_i, for any retailer i, is an *adjusting control variable* for this stock point. Moreover, the order-up-to level at the central warehouse, S_0, constitutes $\mathbf{V}_i$. Let us explain why S_i satisfies conditions 1, 2 and 3 in Axsäter (1990). Condition 1 is fulfilled according to the definition of the model.

To assert that condition 2 is satisfied, we study the system in Axsäter (1990) at an arbitrary point in time. Analyzing the expected costs from this point onwards for an infinite time horizon, it can be shown that the expected future costs are independent of the initial state (the state at this arbitrary point in time). Let us therefore define this initial state so that there are no outstanding orders (i.e. the

stock on hand is S_i at retailer i $\forall$ i and S_0 at the central warehouse). Note that, according to this definition, the initial inventory level at retailer i is dependent on S_i (an increase of S_i by Δ units, increases the initial inventory level at retailer i by Δ units). From this point forward, every unit demanded decreases the inventory level by one unit, and every unit arriving as a replenishment increases the inventory level by one unit.

We can also show that both the future demand and the arrival of future replenishments are independent of S_i. The demand is independent of S_i because the retailer applies complete backordering, which means that all demand is received (and eventually fulfilled) even if there is no stock on hand. Given an initial state as defined above, and for every given sample path of the future demand, the future replenishments are independent of S_i. This is ensured by the following four criteria;

(a) the independency between S_i and orders placed by retailer i,

(b) the independency between S_i and the replenishment process at preceding stock points (the central warehouse),

(c) the independency between S_i and the allocations of orders at preceding stock points, and

(d) the independency between S_i and the shipping of retailer orders to retailer i.

Criterion (a) follows as every customer demand triggers a retailer order, regardless of S_i. Criterion (b) follows from criterion (a) and the fact that the central warehouse replenishment policy only reacts to incoming retailer orders. Criterion (c) follows from criterion (a) and the FCFS allocation policy at the central warehouse. Finally, criterion (d) results from the previous independencies and the deterministic transportation times.

To conclude, recall that according to the definition of the initial state, the initial inventory level shifts with S_i. Thus, for a given initial state and for every given sample path of future demand, an increase or decrease in S_i will simply shift the future inventory level pattern accordingly, verifying condition 2. The independence between S_i and the order placements, allocations and replenishment processes (criteria (a)-(d)) also guarantees that condition 3 is satisfied.

3. Relationship between the Inventory Level Distribution and the Expected Costs

In this section we present the main results regarding the relationships between the expected holding and backorder costs and the inventory level distribution applicable for a stock point i with an *adjusting control variable*, we define:

$\alpha_i(\mathbf{V}_i, S_i)$ = the ready-rate at stock point i $= P\big(IL_i(\mathbf{V}_i, S_i) > 0\big)$

S_i^l = lower bound on S_i, i.e. a (preferably the largest) value of S_i for which $\alpha_i(\mathbf{V}_i, S_i) = 0$.

$IL_i^u(\mathbf{V}_i, S_i)$ = upper bound on the inventory level at stock point i.

From the definition of S_i^l it follows that $IL_i^u(\mathbf{V}_i, S_i^l) = 0$. Using (2) provides the upper bound

$IL_i^u(\mathbf{V}_i, S_i) = S_i - S_i^l$. For the model in Axsäter (1990), a lower bound on S_i is $S_i^l = 0$ and consequently

$IL_i^u(\mathbf{V}_i, S_i) = S_i - S_i^l = S_i$ (independently of $\mathbf{V}_i = \{S_0\}$).

Lemma 1 below is used for obtaining the main result in Theorem 1. But it also leads to Corollaries 1 and 2, below.

Lemma 1. *For stock point* i *with adjusting control variable* S_i, *given* $\mathbf{V}_i = \mathbf{v}_i$

$$C_i(\mathbf{v}_i, S_i) = C_i(\mathbf{v}_i, S_i - 1) + (h_i + \beta_i)\alpha_i(\mathbf{v}_i, S_i) - \beta_i. \qquad (3)$$

Proof: The definition of the upper bound on the inventory level at stock point i gives that $IL_i^u(\mathbf{v}_i, S_i) - 1 = IL_i^u(\mathbf{v}_i, S_i - 1)$. For any discrete stochastic variable Y with a maximum value M

$$\sum_{y=1}^{M} yP(Y = y) = \sum_{z=1}^{M} P(Y \geq z). \qquad (4)$$

Also, for any variable X, $E[X^-] = E[X^+] - E[X]$. These properties, together with (1) and (2) provides

$$
\begin{aligned}
C_i(\mathbf{v}_i, S_i) &= h_i E\big[IL_i^+(\mathbf{v}_i, S_i)\big] + \beta_i E\big[IL_i^-(\mathbf{v}_i, S_i)\big] \\
&= (h_i + \beta_i) \sum_{m=1}^{IL_i^u(\mathbf{v}_i, S_i)} mP\big(IL_i(\mathbf{v}_i, S_i) = m\big) - \beta_i E\big[IL_i(\mathbf{v}_i, S_i)\big] \\
&= (h_i + \beta_i) \sum_{n=1}^{IL_i^u(\mathbf{v}_i, S_i)} P\big(IL_i(\mathbf{v}_i, X_i) \geq n\big) - \beta_i E\big[IL_i(\mathbf{v}_i, X_i)\big] \\
&= (h_i + \beta_i)\left(\sum_{n=2}^{IL_i^u(\mathbf{v}_i, S_i)} P\big(IL_i(\mathbf{v}_i, S_i) \geq n\big) + P\big(IL_i(\mathbf{v}_i, S_i) \geq 1\big) \right) - \beta_i\big(E\big[IL_i(\mathbf{v}_i, S_i - 1)\big] + 1\big) \\
&= (h_i + \beta_i)\left(\sum_{n=1}^{IL_i^u(\mathbf{v}_i, S_i)-1} P\big(IL_i(\mathbf{v}_i, S_i - 1) \geq n\big) + P\big(IL_i(\mathbf{v}_i, S_i) > 0\big) \right) - \beta_i E\big[IL_i(\mathbf{v}_i, S_i - 1)\big] - \beta_i \\
&= (h_i + \beta_i) \sum_{m=1}^{IL_i^u(\mathbf{v}_i, S_i-1)} mP\big(IL_i(\mathbf{v}_i, S_i - 1) = m\big) - \beta_i E\big[IL_i(\mathbf{v}_i, S_i - 1)\big] + (h_i + \beta_i)\alpha_i(\mathbf{v}_i, S_i) - \beta_i \\
&= h_i E\big[IL_i^+(\mathbf{v}_i, S_i - 1)\big] + \beta_i E\big[IL_i^-(\mathbf{v}_i, S_i - 1)\big] + (h_i + \beta_i)\alpha_i(\mathbf{v}_i, S_i) - \beta_i \\
&= C_i(\mathbf{v}_i, S_i - 1) + (h_i + \beta_i)\alpha_i(\mathbf{v}_i, S_i) - \beta_i,
\end{aligned}
$$

proving (3). ∎

Corollary 1. *For stock point* i *with adjusting control variable* S_i, *given* $\mathbf{V}_i = \mathbf{v}_i$, *the inventory costs at stock point* i, $C_i(\mathbf{v}_i, S_i)$, *are convex in* S_i.

Proof: Lemma 1 gives that $C_i(\mathbf{v}_i, S_i) - C_i(\mathbf{v}_i, S_i - 1) = (h_i + \beta_i)\alpha_i(\mathbf{v}_i, S_i) - \beta_i$. This difference increases in S_i as $\alpha_i(\mathbf{v}_i, S_i) = P(IL_i(\mathbf{v}_i, S_i) > 0)$ increases in S_i. ∎

Corollary 2. *For stock point* i *with adjusting control variable* S_i, *for any* $\mathbf{V}_i = \mathbf{v}_i$, *the optimal* S_i *value(s) is (are) the value(s) for which*

$$\alpha_i(\mathbf{v}_i, S_i) \le \frac{\beta_i}{h_i + \beta_i} \le \alpha_i(\mathbf{v}_i, S_i + 1). \tag{5}$$

Proof: From Lemma 1 we know that, given $\mathbf{V}_i = \mathbf{v}_i$, $C_i(\mathbf{v}_i, S_i)$ decreases in S_i, i.e. $C_i(\mathbf{v}_i, S_i) < C_i(\mathbf{v}_i, S_i - 1)$, if $\beta_i/(h_i + \beta_i) < \alpha_i(\mathbf{v}_i, S_i)$. Also, $C_i(\mathbf{v}_i, S_i) = C_i(\mathbf{v}_i, S_i - 1)$ if $\beta_i/(h_i + \beta_i) = \alpha_i(\mathbf{v}_i, S_i)$ and $C_i(\mathbf{v}_i, S_i)$ increases in S_i, i.e. $C_i(\mathbf{v}_i, S_i) > C_i(\mathbf{v}_i, S_i - 1)$, if $\beta_i/(h_i + \beta_i) > \alpha_i(\mathbf{v}_i, S_i)$. (5) follows as $\alpha_i(\mathbf{v}_i, S_i) = P(IL_i(\mathbf{v}_i, S_i) > 0)$ increases in S_i. ∎

Corollary 1 can be used when determining the optimal order-up-to levels under a cost-minimizing strategy. Note also, that for given values of all other variables in the system, also the total system costs are convex in S_i. This is due to the independency between S_i and all other cost components except $C_i(\mathbf{v}_i, S_i)$ (condition 3). Corollary 2 presents a relationship between the ready-rate and the holding and backorder costs in a cost optimal system, which is well-known for single-echelon systems with compound Poisson demand (see, for instance, Axsäter 2006, p. 103). It is here generalized to a group of stock points in multi-echelon systems.

Finally, Theorem 1 below presents the procedure according to which the inventory level distribution can be retrieved from the cost function for this same group of stock points.

Theorem 1. *For stock point* i *with adjusting control variable* S_i, *the probability that the inventory level at stock point* i *is* m, *given* $\mathbf{V}_i = \mathbf{v}_i$, *can be obtained as*

$$P(IL_i(\mathbf{v}_i, S_i) = m) =$$

$$\begin{cases} 0 & , m > IL_i^u(\mathbf{v}_i, S_i) \\ \dfrac{C_i(\mathbf{v}_i, S_i - m + 1) - C_i(\mathbf{v}_i, S_i - m) + \beta_i}{h_i + \beta_i} & , m = IL_i^u(\mathbf{v}_i, S_i) \\ \dfrac{C_i(\mathbf{v}_i, S_i - m + 1) - 2C_i(\mathbf{v}_i, S_i - m) + C_i(\mathbf{v}_i, S_i - m - 1)}{h_i + \beta_i} & , m < IL_i^u(\mathbf{v}_i, S_i) \end{cases} \tag{6}$$

Proof: For $m > IL_i^u(\mathbf{v}_i, S_i)$ the result follows from the definition of $IL_i^u(\mathbf{v}_i, S_i)$. For $m \le IL_i^u(\mathbf{v}_i, S_i)$ Lemma 1 gives that

$$\alpha_i(\mathbf{v}_i, Z) = \frac{C_i(\mathbf{v}_i, Z) - C_i(\mathbf{v}_i, Z - 1) + \beta_i}{h_i + \beta_i}. \tag{7}$$

From (2) follows that

$$\alpha_i\left(\mathbf{v}_i, Z\right) = \alpha_i\left(\mathbf{v}_i, Z-1\right) + P\left(IL_i\left(\mathbf{v}_i, Z\right) = 1\right) \tag{8}$$

and (2) also assures that $P\left(IL_i\left(\mathbf{v}_i, Z\right) = 1\right) = P\left(IL_i\left(\mathbf{v}_i, Z+m-1\right) = m\right)$. Combining these results

provides $P\left(IL_i\left(\mathbf{v}_i, Z+m-1\right) = m\right) = \alpha\left(\mathbf{v}_i, Z\right) - \alpha\left(\mathbf{v}_i, Z-1\right)$. Substituting $S_i = Z + m - 1$ and using (7)

renders the results for $m \le IL_i^u\left(\mathbf{v}_i, S_i\right)$. When m = $IL_i^u\left(\mathbf{v}_i, S_i\right) = S_i - S_i^l$, $\alpha(\mathbf{v}_i, Z - 1) = \alpha(\mathbf{v}_i, S_i - m + 1 -$

$1) = \alpha(\mathbf{v}_i, S_i^l)$ is per definition of S_i^l zero. ∎

The proposed method is applicable also for the central warehouse in some systems. For instance, in

the Axsäter (1990) model, by setting the holding costs and backorder costs at all retailers to zero, the

order-up-to-level at the central warehouse is an *adjusting control variable* of the central warehouse

with $\mathbf{V}_i = \{\}$.

4. Generalizations

Apart from the system in Axsäter (1990), the results in Section 3 hold for many stock points in

different distribution systems. The results can be generalized to different customer demand structures,

to different replenishment policies, and to systems using periodic replenishments.

Following the same reasoning as for the Axsäter (1990) model, it is straightforward to see that

Conditions 1, 2 and 3 specified in Section 2 holds even if the customer demand is compound Poisson

or compound renewal. Note however, that in the case of compound renewal demand, the definition of

the initial state needs some adjustment. For these systems, a state where there are no outstanding

orders may not be reachable (if the customer inter-arrival time cannot take larger values than the

replenishment lead time). Also, for this demand structure, the time since the last order occurrence

need to be included in the state space as it affects the future demand.

Turning our attention to other replenishment policies, the installation stock (R,Q) ordering

policy is defined so that each stock point i places an order of Q_i units whenever the inventory position

(the inventory level + all outstanding orders) is R_i or below R_i. In systems applying (R,Q) ordering

policies, the reorder point of retailer i $\forall$ i, R_i, can fulfill the conditions of an *adjusting control

variable*. For these systems the order quantities of all locations and the reorder points at upstream

locations usually constitute $\mathbf{V}_i$. In order to see this, let the initial state be defined as the state where

there are no outstanding orders and no demands have occurred since the last order placement at each

stock point (i.e. the inventory level = the inventory position = (R_i+Q_i)). A shift of the *adjusting

control variable*, R_i, then clearly shifts the initial inventory level (R_i+Q_i) accordingly. It follows that

for every sample path of the future demand, this shift will be preserved as long as the aforementioned criteria (a)-(d) hold. To assert this, note that given the initial state, the future order placements at retailer i are independent of R_i (criterion (a)). For installation stock policies it is also easy to see that the replenishments at the central warehouse are independent of R_i (criterion (b)). All of the systems cited in Section 1 apply FCFS allocations, ensuring criterion (c). When the shipments are performed with partial deliveries, complete deliveries or a fixed minimum delivery batch and when the shipment times are independent of R_i (e.g. deterministic), criterion (d) is ensured.

For echelon stock (R,Q) policies the inventory position also includes the inventory position of all preceding stock points. For these systems, it is not clear that the warehouse replenishment policy is independent of the reorder point at retailer i, R_i. In order to reach this independency, we replace the control variable R_0 (central warehouse echelon stock reorder point) with $\tilde{R}_0 = R_0 - \sum_i R_i$. For echelon stock (R,Q) systems, the system performance will also be dependent on the initial state of the system. More specifically, it depends on the initial installation stock inventory position at the central warehouse, i_0^0, see Axsäter (1997). However, by considering i_0^0 to be a control variable, as suggested by Axsäter (1997), and by including i_0^0 and $\tilde{R}_0$ in $\mathbf{V}_i$ (instead of R_0), R_i fulfills criteria (a)-(d) (for the same assumptions on allocations and deliveries as above) and constitutes an *adjusting control parameter* for stock point i $\forall i$.

It is also straightforward to apply the results on many systems applying periodic replenishments (with one of the ordering policies discussed above), as long as the replenishments, allocations and deliveries at preceding stock points are performed independently of the *adjusting control variables*. In this case, the replenishment intervals of preceding stock points are usually included in $\mathbf{V}_i$. Consequently, there exists *adjusting control variables* at the retailers in all of the papers referred to in Section 1.

5. Concluding Remarks

This note presents relationships between the inventory level distribution and the costs for a group of systems where there exist parameters fulfilling the conditions of an *adjusting control variable* (see Section 2). In order to find *adjusting control variables* in distribution systems, the replenishments, allocations and deliveries at upstream locations usually need to be independent of the *adjusting control variable*. This occurs for instance when the upstream locations use replenishment policies reacting solely to retailer orders, FCFS allocation policies and partial or complete delivery strategies.

The main result of this note is the derivation of the inventory level distribution from the cost function for systems analyzing the costs directly. All papers for which this result is relevant apply similar cost evaluation methodologies. They study the costs incurring on an arbitrary unit as it travels through the system. Note however that, the results are not dependent on this methodology. Note also that, the presented results are not valid for all papers applying this methodology. One example is the model in Axsäter and Marklund (2008), where *adjusting control variables* are not found at the retailers because the central warehouse uses a replenishment and allocation policy that is dependent on the reorder points at the retailers. There are however a wide range of models for which the results are valid, and hopefully this note can facilitate future research, by identifying *adjusting control variables* in other system settings.

References

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

Axsäter, S. 1993a. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. *Operations Research* **41** 777-785.

Axsäter, S. 1993b. Optimization of order-up-to-S policies in two-echelon inventory systems with periodic review. *Naval Research Logistics* **40** 245-253.

Axsäter , S. 1997. Simple evaluation of echelon stock (R,Q)-policies for two-level inventory systems. *IIE Trans* **29** 661–669.

Axsäter, S. 1998. Evaluation of installation-stock based (R,Q)-policies for two-level inventory systems with Poisson demand. *Operations Research* **46** S135–S145

Axsäter, S. 2000. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand. *Operations Research* **48** 686-696.

Axsäter, S. 2006. Inventory Control. Berlin: Springer

Axsäter, S., J. Marklund. 2008. Optimal position based warehouse ordering in divergent two-echelon inventory systems. *Operations Research* **56** 976–991.

Chen, F., Y.S. Zheng. 1997. One warehouse multi-retailer systems with centralized stock information. *Operations Research* **45** 275-287.

Forsberg, R. 1995. Optimization of order-up-to-S policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research* **81** 143-153.

Forsberg, R. 1997a. Exact evaluation of (R,Q)-policies for two-level inventory systems with Poisson demand. *European Journal of Operational Research* **96** 130-138.

Forsberg, R. 1997b. Evaluation of (R,Q)-policies for two-level inventory systems with generally distributed customer inter-arrival times. *European Journal of Operational Research* **99** 401-411.

Graves, S.C. 1996. A multi-echelon inventory model with fixed replenishment intervals. *Management Science* **42** 1-18.

Marklund, J. 2002. Centralized inventory control in a two-level distribution system with Poisson demand. *Naval Research Logistics* **49** 798-822

Marklund, J. 2011. Inventory Control in Divergent Supply Chains with Time-Based Dispatching and Shipment Consolidation. *Naval Research Logistics* **58** 59-71.

Simon, R.M. 1971. Stationary properties of a two-echelon inventory model for low demand items. *Operations Research* **19** 761-777.

Stenius, O., G. Karaarslan, J. Marklund, A.G. de Kok. 2016. Exact analysis of divergent inventory systems with time-based shipment consolidation and compound Poisson demand. *Operations Research*, forthcoming.

Stenius, O. 2016. Divergent Two-echelon Inventory Systems with Compound Renewal Demand. Working paper Lund University.