



LUND UNIVERSITY

Multi-echelon Inventory Control with Integrated Shipment Decisions

Stenius, Olof

2014

[Link to publication](#)

Citation for published version (APA):

Stenius, O. (2014). *Multi-echelon Inventory Control with Integrated Shipment Decisions*. [Licentiate Thesis, Production Management].

Total number of authors:

1

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Multi-echelon Inventory Control with Integrated Shipment Decisions



LUND
UNIVERSITY

Olof Stenius



Copyright © Olof Stenius

Lund University, Faculty of Engineering LTH
Department of Industrial Management and Logistics
Division of Production Management
978-91-7473-858-2 (printed)
978-91-7473-859-9 (pdf)

Printed in Sweden by Media-Tryck, Lund University
Lund 2014

*To Nasti,
Lykke and Edda*

Abstract

Rising fuel prices and increasing environmental awareness emphasizes the importance of the transportation aspect in logistics. This calls for new improved inventory control methods that consider the effects of shipment strategies in a more realistic manner. This thesis, consisting of an introduction and three scientific papers, studies how shipment decisions can be included in the inventory control of distribution systems. The systems studied in the papers consist of a central warehouse that supplies goods to a number of retailers that face stochastic customer demand.

The first two papers consider a system where shipments from the central warehouse are consolidated to groups of retailers periodically. This means that replenishment orders of one or several items from different retailers are consolidated and dispatched at certain time intervals. By doing so, transportation cost savings can be realized and emissions can be reduced. This is achieved by filling the vehicles or load carriers to a higher extent and by using cheaper and more environmentally friendly, transportation modes.

The first paper explicitly focuses on how to include more realistic transportation costs and emissions. This is done by obtaining the distribution of the size of an arbitrary shipment leaving the central warehouse (directly affected by the shipment frequency). It is thereby easy to evaluate any system where the transportation costs and emissions are dependent on the size of the shipment. The paper also provides a detailed analysis of a system where there is an opportunity to reserve shipment capacity on an intermodal truck-train-truck solution to at least one of the retailer groups. For this system it is shown how to jointly optimize the shipment intervals, the reserved capacities on the intermodal transportation modes and the reorder points in the system. The presented optimization procedure is applicable in three 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 emissions per time unit.

The second paper extends the analysis of a similar time-based shipment consolidation system to handle compound Poisson demand (instead of pure Poisson demand). This system has a simpler transportation cost structure, but the more general demand structure makes the model applicable for a broader array of products. The paper also extends the model to handle fill rate constraints, which further improves the practical applicability. The cost analysis is performed with a new methodology, based on the *nominal inventory position*. This variable is a

helpful tool for analyzing the dynamics of distribution systems. Another system where this tool can be used is studied in the third paper.

In this paper 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. The existing literature predominantly assumes that the available units are shipped immediately and the remaining units are shipped as soon as they arrive to the central warehouse, referred to as *partial delivery*. An alternative is to wait until the entire order is available before dispatching, referred to as *complete delivery*. The paper introduces a cost for splitting the order and evaluates three delivery policies; the PD policy (only partial deliveries are used), the CD policy (only complete deliveries are used), and the state-dependent MSD policy (an optimization between a partial and a complete delivery is performed for each delivery). The MSD policy is proven to perform better than both the PD and the CD policy. In a numerical study it is shown that significant savings can be made by using the MSD policy.

Keywords: Inventory, Multi-echelon, Stochastic, Shipment consolidation, Delivery decisions

Acknowledgements

The work resulting in this thesis was conducted at the division of Production Management in the Department of Industrial Management and Logistics, at Lund University. The Department of Information and Service Economy at Aalto University has generously offered me a place to conduct my research during my stay in Finland. The research has been partly financed by the Vinnova Excellence center NGIL (Next Generation Innovative Logistics) and NordForsk project #25900 Management, design and evaluation of sustainable freight and logistics systems.

First and foremost I want to thank my supervisor Johan Marklund, co-author on two of the papers presented in this thesis. He has been an invaluable help at all stages of the process, particularly at times when I have struggled the most. Johan's experience and solid knowledge of the field has helped me find interesting research paths. He has kept my feet on the ground when I have wanted to leap to ten cliffs in different directions before settling from the previous leap. He has also patiently given me both general and detailed comments on my drafts. These thorough comments have improved this thesis tremendously, and helped me develop as a researcher.

My colleague and co-author of the third paper in this thesis, Christian Howard, has also played an absolutely crucial role in the completion of this thesis. Christian has been supporting me on all levels throughout the process and has become one of my closest friends. Thank you for listening and providing sane feedback on all of my brilliant (and sometimes less brilliant) ideas. Thank you also for taking care of me, especially during my stays in Lund.

I would like to express the deepest appreciation to Sven Axsäter, my assistant supervisor and co-author of the first paper. Sven is a great inspiration to me and has generously offered guidance and expertise on all papers. Thank you also Gönül Karaarslan and Ton de Kok for the collaboration resulting in paper II. The process of writing that paper was critical for my development as a researcher. I would also like to thank all of my wonderful colleagues at the department: Peter Berling, Fredrik Eng Larsson, Lina Johansson, Fredrik Olsson, Birgitta Olvegart, Pasi Porkka and everybody else.

Last but not least I would like to thank all of my family and friends for the support. The deepest gratitude goes to my wife, Nasti, who has the ability to always bring out the best in me. Thank you.

Lund 2014-01-10
Olle Stenius

Contents

ABSTRACT	I
ACKNOWLEDGEMENTS	III
CONTENTS	V
1. LIST OF PAPERS	1
2. INTRODUCTION	3
2.1 Research Objectives	4
2.2 Research Methodology	5
2.3 Brief Overview of Inventory Control Theory	6
2.3.1 <i>Structure</i>	7
2.3.2 <i>Lead Times</i>	8
2.3.3 <i>Customer Demand and Review Periods</i>	9
2.3.4 <i>Performance Measures</i>	10
2.3.5 <i>Replenishment Policies</i>	11
2.4 Summary of Papers	12
2.4.1 <i>Paper I – Sustainable Multi-echelon Inventory Control with Shipment Consolidation and Volume Dependent Freight Costs</i>	13
2.4.2 <i>Paper II – Exact Analysis of Divergent Inventory Systems with Time-Based Shipment Consolidation and Compound Poisson Demand</i>	14
2.4.3 <i>Paper III – Partial or Complete Deliveries in Two-echelon Inventory Systems?</i>	16
2.5 Contributions and Future Research	18
REFERENCES	21

1. List of Papers

- **Paper I – Sustainable Multi-echelon Inventory Control with Shipment Consolidation and Volume Dependent Freight Costs**
Axsäter, S., J. Marklund, O. Stenius
Department of Industrial Management and Logistics, Lund University
(2013) 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
(2013) Under revision
- **Paper III – Partial or Complete Deliveries in Two-echelon Inventory Systems?**
Howard, C., O. Stenius
Department of Industrial Management and Logistics, Lund University
(2013) Submitted

2. Introduction

Efficient control of distribution systems is a core activity in many companies. These distribution systems often consist of a central warehouse and a number of local warehouses (retailers). The central warehouse creates economies of scale by ordering larger quantities of some products from an outside supplier or is located in conjunction to a production facility. The retailers are located close to the customers to provide products by selling them from the shelf or by ensuring sufficiently short service times when delivering to the customers. The control of these systems is focusing on ensuring high customer satisfaction (high fill rates or low shortage costs) while keeping costs of holding inventory and procuring replenishments to a minimum. This task is complicated by the fact that future demand is unknown.

These problems are common in practice, and naturally, inventory policies for distribution systems is a well-studied field in academia (see, for example, Axsäter, 2006). The research in this thesis presents new approaches for exact analysis of distribution systems, which allow for more general and realistic replenishment and transportation cost structures.

In recent years, the shipment aspect of logistics planning has gain increased attention due to increasing fuel prices and growing awareness of environmental issues within companies and their customers. For instance, the transportation sector was responsible for over 20 % of the greenhouse gases emitted in Europe 2007 (European Commission, 2011), and the share of emissions caused by goods transports is increasing (European Commission, 2009). This creates a need for more sophisticated control policies for distribution systems, where both the transportation costs and the emissions during transports are considered. This can be achieved by assuming more realistic replenishment costs when deciding on replenishments as studied in paper III of this thesis. Another way to reduce transportation costs and emissions is to consolidate shipments, i.e. coordinating shipping and transportation activities across different products and/or different retailers. This issue is studied in Paper I and II of the thesis.

2.1 Research Objectives

The research objective of this thesis can be summarized as:

To develop new exact inventory control methods for distribution systems with stochastic demand, with a focus on models that incorporate shipment decisions.

This research objective might need some further clarification. By “develop... inventory control methods” we mean that for a reasonable replenishment policy, a method to determine the costs and optimizing the system parameters is presented. By exact we mean that the exact expected long-run costs under given assumptions are obtained and that the optimal policy given these assumptions is found. The statement “...incorporating shipment decisions” specifically means that we study two ways of incorporating shipment decisions in the inventory control models.

First, we study how a distribution system can be controlled under time-based shipment consolidation. This means that shipments from the central warehouse leaves (with one or multiple products) to groups of retailers periodically. We develop models that determine how often shipments should be made and how much stock should be kept at each location under different transportation cost structures. We also include the emissions during transportations into the decision variables of the problem. For instance, we study how the system should be controlled if there is a restriction on how much emissions the transportation on average can produce. These issues are examined in Paper I and Paper II of the thesis.

Second, we study a distribution system with batch ordering, and investigate when partial or complete deliveries from the central warehouse should be used. To explain this, consider the following example: A retailer orders a batch of five units from the central warehouse. The warehouse only has three units available, but there is a replenishment arriving to the central warehouse in two days. The warehouse now needs to decide whether she should split the order and ship the three available units immediately (partial delivery) or whether she should wait until the entire order is available in two days (complete delivery). By introducing a cost for splitting the order, this delivery decision is optimized. The contributions include derivation of the exact average cost for the system, when this policy is used, and an optimization procedure for the reorder points at each location. The performance of the new policy is compared to systems where only partial or only complete deliveries are used. This research is presented in paper III of the thesis.

2.2 Research Methodology

As stated in the previous section, the research objective is to develop exact inventory control methods. The tools chosen to reach this goal are based on applied mathematical modeling. Modeling of stochastic inventory systems are typically based on 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.

The modeling process of these systems can be divided into three steps, which generally have to be repeated iteratively (e.g., Hillier and Lieberman, 2010, and Axsäter and Marklund, 2010). The first step is to build the model, by deciding on the structure of the system and the assumptions made. In this step the right level of detail has to be ensured. It is important to capture the main characteristics of the system, but usually some simplifying assumptions of the real system have to be made in order to reach any useful results. The model can either aim at being general (applicable for various different companies and cases) or adjusted to a specific environment. The research in this thesis focuses on general models, with the aim of being applicable on many distribution systems. However, when applied, these general models usually need to be adjusted further to fit the specific cases.

The second step is to generate a solution. For this step, the tools from some of the mathematical fields listed above are applied and developed. The methods used in this thesis are mainly based on probability theory. Because of the stochastic demand in the models, the main mathematical challenges lies in modeling the dynamics of the system as the demand occurs, thus finding the relationships between different stochastic variables (such as demand and stock on hand). For distribution systems these relationships tend to generate quite complex problems. When these relationships are solved, the expected costs for a given policy can be obtained. Optimization methods are then used and applied in order to find the optimal policy.

The third step is to validate the results. This step is needed to ensure that the generated solution is correct. If the validation fails, one might have to return to previous steps to regenerate a solution or to modify the model.

The models in this thesis are evaluated exactly. This means that after the model assumptions regarding the system operators are set, no approximations are introduced to facilitate the cost evaluation and optimization of the model. This exact solution procedure is not an obvious choice for complex Operations Research problems. The solution achieved is of course the best possible, but in practical applications the exact evaluations can sometimes be too computationally cumbersome to implement. Especially in systems with a large amount of different

products or where the demands of the products are very large. In these situations a heuristic ensuring a reasonably good solution with shorter computation times might be preferable. The motivation to focus on the exact solutions, apart from the fact that they create stable implementable solutions for smaller systems, is that future heuristics can be based on the exact solution 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 and optimality of different policies.

Apart from exact and heuristic solutions, another possible way to model inventory systems is by simulation. Simulations have the advantage of being flexible in the sense that it is possible to evaluate complex systems and policies. There are however a few drawbacks to this methodology. Firstly, optimization via simulation search is often very time consuming. One simulation run will typically provide the result for one set of parameters only. Also, the simulation run times tend to be quite long to get good solutions and the parameter sets grow rapidly as the systems get 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 validation of the results.

2.3 Brief Overview of Inventory Control Theory

This section introduces the reader to the basics of inventory control theory. Theories and concepts are explained, with an aim to facilitate the reading of the papers in the thesis. As the models in this thesis are based on stochastic demand (future demand is unknown), 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. This problem studies a newsvendor, who wishes to optimize how many 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, but avoid having lots of excess papers unsold at the end of the day. This relatively simple inventory control problem, is a so called single period problem, and only deals with the question of how much material to order? However, when you allow for material to be stored and sold later, you usually

also deal with the question of when to order. In the remainder of this section we will describe features that define different inventory control problems, with a focus on features of the problems studied in this thesis.

2.3.1 Structure

The structure, or the topology, of the problem describes how many stock points that are controlled in the system, and how they are connected, i.e. from where each stock point receives its replenishments and how material flow 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 distribution 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 unit, and local warehouses closer to the different markets. An example of a two-echelon distribution system is illustrated in Figure 3.

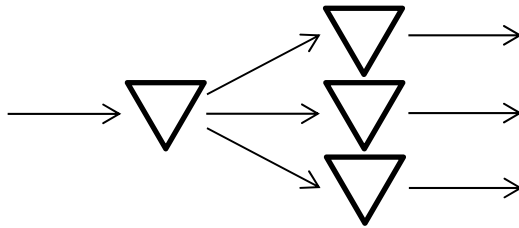


Figure 3. Example of a two-echelon distribution system

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.

Another structural feature that has gained increased attention during the last decades is the concept of lateral transshipments. This means that different stock points at the same echelon (stage) shares inventories and transships goods between them as needed. Consider, for instance, a situation when a retailer without stock on hand receives a customer order of one unit. In this case, it may be more efficient to request this unit from another retailer, located nearby, instead of from the warehouse the retailer normally replenishes from, as this warehouse often is located further away from the retailer. This issue is excluded from the systems in this thesis, but might be interesting subjects for future research. For a recent overview on literature of lateral transshipments in distribution systems 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 usually also includes the time for order placement, picking, loading and receiving activities. In multi-echelon systems it generally also include the time spent waiting for products or components to become available at stock points upstream. Often, in practice, the lead time varies, but if the variation is small, it is common to assume that the lead time is constant in the models. This facilitates the analysis of more complex problems. The lead time variation is also excluded from the systems studied in this thesis apart from the variation caused by stock-outs at the preceding stock

points, which in many multi-echelon systems causes the majority of the lead time variation.

2.3.3 Customer Demand and Review Periods

The representation of the customer demand differs dependent 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 majority of the inventory control problems can be grouped in either 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. For an overview of the literature on inventory control in distribution systems with periodic review, see, for example, Axsäter (2003).

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 made continuously. In continuous review systems the demand is often characterized by a process, which means that the probabilities of when customers will arrive and how much they will order are known¹. The most commonly used process is the Poisson process. When customer demand follows a Poisson process, it can be interpreted as a system where each customer arrives independently of the others and that each customer orders only one unit. The Poisson process is characterized by the arrival intensity of the customers (amount of customers per time unit). Due to the independency, the Poisson process has several analytical advantages, but it is not always a good representation of the actual demand pattern. One problem with the Poisson process is that the demand during any given period always has a variance that is equal to the mean (the variance to mean ratio is equal to one). Clearly, this may not always be a good representation of the actual variance of the demand, seen in the system. However, in many systems dealing with, for instance, spare parts, the Poisson process is a good approximation of the real demand process. The demand is assumed to follow a Poisson process in paper I and III of this thesis.

¹ In some simpler systems the distribution of the demand during a replenishment lead time is sufficient to analyze the system.

For a more flexible demand representation, the compound Poisson process can be used. Here, the customers still arrive independently. 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 compound Poisson process can handle any variance to mean ratios 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. For even more general demand structures, compound renewal demand processes can be used. Here the times between two arriving customers can follow any distribution (in the Poisson and compound Poisson processes they always follow an exponential distribution). These demand structures are very flexible, but tend to be difficult to analyze when applied to complex multi-echelon systems.

2.3.4 Performance Measures

The performance measures characterize what the aim of the inventory control is. In the example with the newsvendor it is commonly assumed that there is a fixed purchasing price per paper and a fixed selling price per paper. The performance measure is thus the expected profit of the newsvendor. This is not unusual for inventory control problems, the aim often being 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, replenishment costs and backorder costs. The holding cost is the costs for storing the product (e.g. opportunity cost for tied-up capital, material handling, storage, damage and obsolescence, insurance, and taxes) and is often assumed to incur proportionally to the amount of units stored and the time they are stored. The replenishment cost (or order/setup cost) is the cost associated with each replenishment request. These fixed costs can occur, for instance, in production, during transportation or in administration of the order. The replenishment costs are 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. In different companies these costs can vary a lot. 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 while the product is backordered. In the papers in this thesis it is assumed that the customers will wait for the product. When this happens, costs are incurring for the company. These costs include administrative costs, possible price reductions offered to the customer and the trade mark damage caused by the shortage. Often these costs are increasing the longer the customer has to wait. It is therefore not

unusual to assume that the backorder cost is proportional to the amount of products and the time the products are backordered. 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. One commonly used service level is the fill rate, i.e. the proportion of demand that can be satisfied directly from stock. If service level constraints are used, the aim of the inventory control is to minimize the (holding and replenishment) costs while assuring that the service constraints are met. Papers I, and III in this thesis are based on backordering costs, while in Paper II, either backordering costs or fill rate constraints can be used.

2.3.5 Replenishment Policies

The replenishment policy states 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 continuous time single-echelon systems 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 continuous review models where customers orders one unit at a time, the (s,S) policy is equivalent to an (R,Q) policy (with $R=s$ and $R+Q=S$). With the (R,Q) policy an order of Q units is placed as soon as the inventory position drops to, or below, the reorder point, R . This policy has the advantage that orders always are placed in multiples of Q , which can be a beneficial feature in practice, due to, for instance, packaging sizes.

In serial systems where there are replenishment 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 (one-for-one replenishments) 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 and Chen, 2000), 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 has thus focused on determining the costs (exactly or by heuristics) and optimizing system parameters under reasonable replenishment policies and comparing these different replenishment policies.

The terms replenishment policy and ordering policy is often used synonymously. In this thesis however, the replenishment policy consist 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 batching orders decrease. However, the incentives for batching 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 location share all his demand information to the upstream location and let the upstream location decide on the batching activities. The upstream location has more information and can for instance batch units of different products and from 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 available units to satisfy all requests, the upstream stock point is faced with an allocation decision. According to the First-Come-First-Served (FCFS) allocation policy, the downstream facility that placed the order first will receive their replenishments first. This assumption facilitates the analysis and can also be seen as a “fair” policy to use in practice, but it is not optimal. FCFS is the dominant allocation policy used in the continuous review literature, and also used in some of the periodic review literature (e.g. Axsäter 1993, Graves, 1996, Shang and Zhou, 2012). 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 made, but on average these savings are less than 3 % in their numerical studies, which indicate that FCFS is often reasonable to use.

2.4 Summary of Papers

This section summarizes the most relevant aspects of the three papers in the thesis. 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 of each paper will be highlighted in Section 2.5.

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

As indicated earlier, the costs of placing orders are diminishing, while there are significant costs associated with dispatching orders, such as costs for transportation, picking, receiving administration etc. Increased fuel prices and environmental awareness has emphasized the relevance of explicitly considering transportation costs and emissions during transportations in inventory control decisions. In this paper, we therefore study a distribution system with consolidated shipments from the central warehouse, where realistic transportation costs and emissions are included in the model.

The system consists of a central warehouse supplying a number of retailers that each faces independent Poisson demand. The retailers use (S-1,S) ordering policies (an order is placed immediately when a customer demand occurs) to transfer the demand information to the warehouse. Batching of the orders is performed at the warehouse, where the shipments are consolidated and dispatched 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). The warehouse replenishes from an outside supplier with an (R,Q) policy. The model can handle situations where consolidation is performed over a single item (product) or jointly over multiple items. Figure 4 illustrates an example of the structure of the model where there are three retailers belonging to two retailer groups.

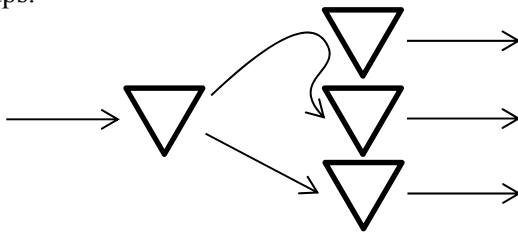


Figure 4. Example of the structure in Paper I and II.

For this system, Marklund (2011) presents a fast method for exact evaluation of the expected inventory holding and backorder costs. In this paper, the model is extended by including more realistic transportation costs and emissions. This is done by obtaining the distribution of the amount of units that are dispatched on an arbitrary shipment. When this distribution is known, the model can handle all transportation structures, where the transportation costs and emissions are dependent on the shipment size.

An example of such a transportation structure is analyzed in detail in the paper. The setup is motivated by discussions with several industry partners that

are interested in (or have already implemented) train transports from warehouses in central Europe to retailers in Scandinavia. Purchasing these train transports (or in fact intermodal truck-train-truck transports) requires reservation of capacity on the trains in advance. When dispatched, this reserved capacity (w_k) is used to maximum extent and excess units are shipped with trucks directly from the warehouse to the retailers. In the model, there are fixed costs associated with each shipment leaving, depending on how much capacity that have been reserved. There are also fixed costs for each truck in use and costs per unit for transporting with both transportation modes. The transport emissions in the system are modeled by equivalent parameters.

The paper presents a method to evaluate the exact expected costs as well as 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 it is illustrated how the expected total cost $TC^*(\theta)$ can increase as the θ value decreases when optimizing scenario (iii) (for the details of the study, see the paper). In the example, emission cuts of 16.5% can be achieved with a cost increase of only 1.5%, while larger emission cuts will be relatively expensive (see Figure 5).

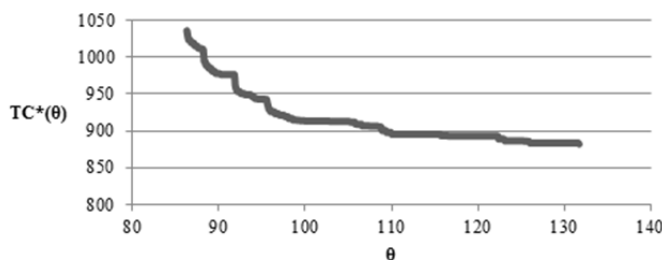


Figure 5. Expected total cost for varying θ values.

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

This paper investigates a system that is similar to the one considered in Paper I. Again, there is a central warehouse supplying retailers and shipments are consolidated and dispatched to groups of retailers periodically (see Figure 4, above). The main distinction from Marklund (2011) and Paper I is that the

demand is now assumed to follow a compound Poisson process (note that the Poisson process assumed in Marklund, 2011 and Paper I 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 also 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 a fixed costs for each scheduled shipment (independent of volume) and costs per unit for each transported unit. These assumptions are realistic in many cases where the transportations are bought from an outside third or fourth party logistics provider. This assumption simplifies the analysis of the shipment costs. 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.

The complicating matter in the analysis of inventory levels in distribution systems is that the performance of the central warehouse affects the retailers. The moment a retailer will receive an order depends on whether there are units available at the central warehouse or not. In this paper this issue is solved by obtaining the exact distribution of the amount of backorders at the central warehouse destined to a specific retailer (when a shipment is leaving to this retailer). Through this distribution it is fairly straightforward to generate the distribution of the inventory levels at the retailers. This analysis of the backorders at the central warehouse is performed using a new methodology, which can be used to solve also other types of inventory control problems in distribution systems (for instance, a version of the methodology is used in Paper III).

The methodology is based on the *nominal inventory position*, which is a tool that facilitates the analysis of the backorders at an arbitrary point in time t_0 . It is a well-known result that the inventory level at time t_0 (= the stock on hand – the backorders) is equal to the inventory position (= inventory level + outstanding orders) a replenishment lead time, L_0 , earlier minus the demand between these points in time. The *nominal inventory position* is defined as the inventory position at $t_0 - L_0$ minus the demand that have already occurred at the central warehouse since $t_0 - L_0$. Thereby, the *nominal inventory position* 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 . The analysis of the backorders at the central warehouse

with the aid of the *nominal inventory position* constitutes the backbone of the analysis in this paper.

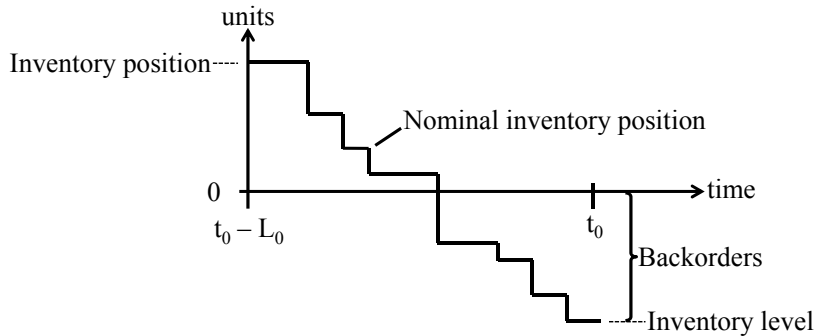


Figure 6. A sample path of the nominal inventory position.

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 proposed analytical method for cost evaluation and the optimization procedure are applicable in both single- and multi-item systems.

In a numerical study, it is investigated how the control parameters are influenced by parameter changes in the system. The study shows that the shipment intervals are primarily used to control against changes in the shipment cost, but it also indicates that shorter shipment intervals should be used in systems where the variance to mean ratio and the backorder costs are high.

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

This paper studies a distribution system with 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). 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 replenishment 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 once, i.e. a *complete delivery*.

This paper introduces this cost of splitting an order, θ_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.

The existing literature predominantly assumes PD policies. In this 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, θ_i , is 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.

For all three policies, the exact expected costs in the system are evaluated. Because the MSD policy contains a state-dependent decision (dependent on when outlying 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 slightly 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 of different delivery decision occurrences. 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 the paper for details of the study). The study consists of 32 problems, which are all optimized under all three policies. For these problems, the expected costs under the PD policy are on average 5.8 % higher than under the MSD policy and the costs under the CD policy is 5.9 % higher than under the MSD policy. The maximum increase of using the PD policy over 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 very high. The study also indicates that when the costs for splitting orders, θ_i , 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 seen previously in the literature (see e.g. Axsäter 2003). The majority of the inventory will, however, still be kept at the retailers.

2.5 Contributions and Future Research

This section highlights the contributions of the research in this thesis and provides some interesting directions for further research.

Paper I introduces a new way to analyze distribution systems. To the best of our knowledge this paper is the first model studying a stochastic multi-echelon system that includes size 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. This analysis is performed by a binomial disaggregation technique. Similar methodologies have been used by, for instance, Simon (1971) and Graves (1985) for a different problem. They use binomial disaggregation in a distribution system with $(S-1,S)$ ordering policies, Poisson demand and no shipment consolidation to receive the distribution of central warehouse backorders destined to a specific retailer. This is however the first time it is used to determine the distribution of the shipment size. 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.

Interesting directions of further research include more thorough evaluation of other transportation cost structures. Extensions of the current model to handle compound Poisson demand would also be relevant.

Paper II generalizes the model of Marklund (2011) in three ways. Firstly, it handles compound Poisson demand instead of Poisson demand. It can therefore be applied to any system where the variance to mean ratio is larger than or equal to one. Secondly, it extends the analysis to handle fill rate constraints apart from 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 nominal inventory position, can be seen as a contribution itself, as it has already proven to be useful in other contexts.

Future research directions include a variety of models where variations of this methodology are applied. Paper III provides one such work, but we also believe that the methodology can be used for, for instance, models with compound renewal demand and/or more than two echelons. Another interesting path for future research would be to provide new, fast heuristics to analyze this system, as the computational times with the proposed exact method escalate when the number of retailers, demanded units and demand sizes increase. Finally, it would be interesting to compare the results of this time-based shipment consolidation policy with other shipment consolidation policies, for instance quantity-based shipment consolidation. With such a policy, orders are consolidated until a certain quantity is reached. From a sustainability perspective, this methodology would be interesting as it could assure full utilization of the load carriers on the transportation modes.

Paper III introduces a more realistic cost structure to distribution systems with batch ordering. By introducing the 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 MSD policy, which has a performance guarantee over the other two. The main analytical contribution of this paper is the derivation of the exact expected costs under this policy. In this analysis, new useful features of the nominal inventory position are observed and used. It is our strong belief that these features will prove to be useful also in other problem contexts.

The numerical study indicates that substantial savings can be made by using our policy compared to the simpler ones. 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. This is one interesting direction for future research. It would also be interesting to extend the analysis to handle compound Poisson demand.

References

- Axsäter, S. 1993. 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. 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. Springer, New York.
- Axsäter, S., J. Marklund. 2010. *Decision sciences*, Encyclopedia of library and information sciences, third edition **1:1** 1450-1457.
- Chen, F. 2000. Optimal policies for multi-echelon inventory problems with batch ordering. *Operations Research* **48** 376-389.
- Clark, A. J., H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Science* **6** 475-490.
- European Commission. 2009. European energy and transport trends to 2030 - Update 2009. Tech.rept. European Commission. Directorate-General for Energy and Transport.
- European Commission. 2011. Reducing emissions from transport. http://ec.europa.eu/clima/policies/transport/index_en.htm, accessed 08/16/2013.
- Federgruen, A., P. Zipkin. 1984. Computational issues in an infinite-horizon, multi-echelon inventory model. *Operations research* **32** 818-836.
- 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.

- 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.
- Marklund, J. 2011. Inventory control in divergent supply chains with time based dispatching and shipment consolidation. *Naval Research Logistics* **58** 59-71.
- 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.H., S.X. Zhou. 2012. Optimizing Replenishment Intervals for Two-echelon Distribution Systems with Stochastic Demand. Working paper, Fuqua School of Business, Duke University, North Carolina, USA.
- 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.
- 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.

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

Sven Axsäter • Johan Marklund • Olof Stenius

Department of Industrial Management and Logistics, Lund University

svan.axsater@iml.lth.se • johan.marklund@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. 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 realistic volume dependent freight cost structures and emissions to be included in the model. Bearing in mind the generality of these results, we focus on a particular model setting that emphasizes the model's usefulness from a sustainability perspective. The model setting is inspired by discussions with industry and involves the possibility to reserve transportation capacity on an intermodal shuttle train solution, and to use truck transports available on demand. 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 a cost of emissions. 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.

Keywords: Inventory, Multi-echelon, Stochastic, Shipment consolidation, Shipment size, Sustainability

1. Introduction

Higher fuel prices and increasing environmental concerns drive a growing interest by companies and their stakeholders to strive for more sustainable freight transportation and distribution systems. An important aspect of this challenge is to reduce transportation emissions while minimizing the total inventory and transportation costs. As greener transportation solutions often tend to be cheaper but slower, a key issue is to balance increasing inventory costs associated with longer lead times against potential savings in transportation costs and reduced emissions. Clearly, considering stochastic multi-echelon inventory decisions jointly with transportation decisions offers a promising course of action for finding efficient solutions. However, the existence and use of such models, so far, is scarce. This paper aims to help remedy this by presenting a model for a distribution system with a central warehouse that replenishes the stock of N non-identical retailers (or local warehouses) using a time based shipment consolidation policy with volume dependent freight costs and emissions.

The basic model settings are similar to those in Marklund (2011), with two important differences; inclusion of the volume dependent freight costs and emissions. These added features constitute major extensions that make the current model more general and realistic with respect to the transportation decisions. A key contribution that enables these extensions 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.

Bearing in mind the generality of our underlying results, we focus the model formulation on a particular context that emphasizes the model's usefulness from a sustainability perspective. The model encompasses an intermodal transportation option by train in combination with truck transports. It is motivated by discussions with several Swedish companies having central warehouses (either owned by themselves or by their suppliers) in central parts of Europe. Spurred by ambitions to reduce costs and transportation emissions they investigate (or have already embarked on) solutions where goods from the central warehouses are shipped to Sweden by train. More precisely, a transport provider offers a periodic intermodal shipment solution, where by far the longest traveled distance will be on the train. In practice they offer an opportunity to use truck-train-truck solutions, where (shuttle) trains leave periodically from logistics hubs in central Europe according to predefined schedules. Procurement of these intermodal transportation services requires that capacity on the train is reserved in advance for a given contract period.

Looking closer at our model set up, it assumes a centralized system in the sense that 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. Demand at each retailer is assumed to be Poisson and the warehouse allocates the units to the different retailers according to a First-Come-First-Served (FCFS) policy. 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 shipment interval for the group in

question. At each dispatching opportunity the warehouse ships all demanded units that are available. The intermodal transportation capacity is used to the maximum extent, and units exceeding the reserved capacity are shipped with an alternative transportation mode (e.g. trucks) offering the same transportation time, but typically with higher variable costs and emissions per unit. These assumptions stem from discussions with industry representatives. The replenishments to the central warehouse is performed with a continuous review (R,Q)-policy from an outside supplier/manufacturer.

Marklund (2011) presents an exact recursive procedure for obtaining the expected inventory holding and backorder costs for the considered system. However, with regards to the shipment cost structure this model is quite restrictive, assuming a single mode of transportation, fixed costs for every scheduled shipment and no emissions. The contributions of our present work include the 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 contribute with a joint optimization procedure of the shipment intervals (to each retailer group), the reserved intermodal capacity (to each retailer group), and the reorder levels (at all stock points), and illustrate how the transportation emissions can be considered in this optimization. The presented analysis focuses on the single-item case, but extensions to multi-item settings are also provided.

The remainder of this section is devoted to an overview of the related literature. Section 2 provides the problem 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. Section 5 provides an illustrative example. Section 6 extends the model to a multi-item setting, and discusses some other generalizations. Section 7 summarizes and concludes.

1.1 Related literature

As indicated above, our work is closely related to Marklund (2011) which analyses the same type of inventory system as we do but under a much more restrictive transportation cost structure, which does not allow for inclusion of realistic transport emissions. A key assumption in this model is that there is a fixed cost for each scheduled shipment leaving for each retailer group, regardless of the quantity shipped on that occasion (even if no units will be shipped). 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 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 papers are 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. (2013) generalize the work of Marklund (2011) to a compound Poisson setting. 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. The cross-docking facility orders jointly for a group of customers, but does not hold any stock.

Apart from these articles, our current work is related to the literature on 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 one-for-one replenishments. Generalizations to compound Poisson demand and/or batch ordering are provided in Axsäter (1993a, 1997, 1998, 2000), Forsberg (1995, 1997) and Chen and Zheng (1997). The same research stream also encompass papers investigating more elaborate replenishment and delivery policies, including

Marklund (2002, 2006), Moinzadeh (2002), Axsäter and Marklund (2008) and Howard and Stenius (2013).

Because of the periodic shipments leaving from the central warehouse, our work is connected to the literature studying periodic review multi-echelon inventory control. In contrast to what is assumed in our work, the main body of the periodic review literature does not make use of real-time information (see, for instance, Federgruen and Zipkin, 1984 a,b, Federgruen, 1993, Axsäter, 1993b, Graves, 1996, Verrijdt and de Kok, 1996, Houtum et al., 1996, Heijden et al., 1997, Cachon, 1999, Cachon and Fisher, 2000, Axsäter et al., 2002, Özer, 2003, Dogru, 2006, Chu and Shen, 2010, Shang and Zhou 2012, Marklund and Rosling, 2012, and references therein). Inventory levels are typically observed only when replenishments may be placed. However, Graves (1996), Axsäter (1993b) and Shang and Zhou (2012) assume virtual FCFS allocation of orders under Poisson demand arrivals. This essentially means that demand is monitored continuously although orders are triggered periodically. Graves (1996) and Axsäter (1993b) consider fixed review periods while Shang and Zhou (2012) show how to optimize the reorder intervals. The latter is related to the optimization of the shipment intervals in our study. Compared to our present work they use simpler replenishment cost structures (independent of the shipment size) and do not consider transportation emissions.

Clearly, our work is also associated with the shipment consolidation literature focusing on single-echelon settings (e.g. Çetinkaya and Lee, 2000, Axsäter, 2001, Çetinkaya and Bookbinder, 2003, Chen et al., 2005, Çetinkaya et al., 2006, Çetinkaya et al., 2008, Mutlu and Çetinkaya, 2010, 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. Especially the work by Çetinkaya and Bookbinder (2003), Mutlu and Çetinkaya (2010) and

Kaya et al. (2012) are connected to our work as they assume volume dependent dispatching costs. Apart from the fact that we study a multi-echelon model, a key difference between our work and this stream of literature is that we explicitly consider emissions in the optimization.

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 reverse and closed-loop logistics or on waste management (see, for example, Corbett and Kleindorfer, 2001a, 2001b, Kleindorfer et al. 2005, Corbett and Klassen, 2006, Srivastava, 2007, and Dekker et al., 2012, for overviews). Lately, the interest in models where transport emissions are considered has increased. One stream of research investigates lot sizing in single-echelon and two-echelon systems with deterministic demand, under different kinds of environmental regulations, see, for example, Hua et al. (2011), Bouchery et al. (2012), Benjaafar et al. (2013), and, Jaber et al. (2013). In stochastic demand models, Berling and Eng-Larsson (2013) investigate different contracts between a transport provider and a buyer with a warehouse. They study two different transportation modes, one of which requires reservation of capacity. Hoen et al. (2012) explore different transportation alternatives with different transit times in a single-echelon newsboy setting. Recent work in the field of sustainable supply chains also includes Caro et al. (2013), who study allocation of greenhouse gas emissions between different firms, when the firms are jointly responsible for the emissions. Hoen et al. (2013), study the choice of transportation modes and prices to a portfolio of customers, with the objective to stay below a carbon emissions cap. To the best of our knowledge, our present work is the first to consider volume dependent shipment costs and transport emissions in a stochastic multi-echelon inventory setting.

2. Model formulation

The N non-identical retailers face independent Poisson demand, and are replenished from a central warehouse with access to real-time inventory information about the entire system. The central warehouse replenishes its stock from an outside supplier with constant leadtimes. Complete backordering and FCFS allocation is assumed at all stock points.

The retailers are divided into K retailer groups ($K \leq N$) 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. More precisely, the central warehouse dispatches a shipment to all retailers within retailer group k ($1 \leq k \leq K$) 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 some type of vehicle routing method (see, for example, Toth and Vigos, 2001 for an overview). The key characteristic is that replenishments to retailers in the same group can be consolidated, offering a potential for reducing transportation costs and emissions. 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 and constraints beyond the scope of our present model and is left for future research.

To each of the retailer groups, shipments can be made either using a capacitated intermodal transportation option (where capacity is reserved in advance), or an alternative option with unlimited capacity. Both options are offering the same transportation times. It is assumed that whenever the shipment quantity to a specific retailer group exceeds the reserved intermodal capacity, the excess units are shipped using the alternative transportation option.

The set of possible capacity reservations for the intermodal 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., $0 \in W_k \forall k=1,2,\dots,K$), and only use the alternative mode of transportation. In practice, the capacity reservation opportunities to a specific retailer group typically include the possibilities to reserve nothing, an integer multiple of whole trains or an integer multiple of freight cars.

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 . We assume that the order quantity Q_0 is determined by the outside supplier, taking production set up costs etc. into consideration. As mentioned earlier we assume a centralized system where 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 orders are batched and 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.

The replenishment lead time for orders placed by the central warehouse, L_0 , is fixed. The transportation time 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 mode used. However, the replenishment lead time to retailer i is stochastic due to the shipment intervals and potential stock outs at the central warehouse.

Focusing on the inventory process at the central warehouse, the FCFS allocation implies that a unit will be reserved for a specific retailer at the moment this retailer experiences a customer demand. This reserved unit might 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 leaving to the retailer in question. In case the reserved unit is not yet available, it will be backordered at the central warehouse. In this case, the unit becomes *qualified* for shipment (and the backorder cleared) at the moment it is delivered to the central warehouse from the outside supplier. After becoming qualified for shipment, the unit will be sent to the retailer in question on the next shipment.

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 the central warehouse 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} = the vector of the replenishment intervals to all retailer groups = $\{T_1, \dots, T_K\}$, $T_k = n_k \cdot T_{\min} \quad \forall k=1, 2, \dots, K$, where n_k is an integer greater than zero.

\mathbf{w} = the vector of capacity reservation quantities to all retailer groups = $\{w_1, \dots, w_K\}$, $w_k \in W_k \quad \forall k=1, 2, \dots, K$.

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

λ_i = the expected demand per time unit at retailer i .

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

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

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

b_i = backorder cost per unit and time unit at retailer i , $i = 1, 2, \dots, N$

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 the expected inventory holding and backorder cost per time unit, TIC , is independent of the reserved capacity \mathbf{w} . The reason is that all units qualified for shipment at any given shipment opportunity will be shipped either using the capacitated intermodal transportation option or the alternative option, and their transportation times are the same. Also note that the expected shipment cost per time unit, TSC , is independent of the base-stock levels, \mathbf{S} , (see Section 3 for explanation), and recall that Q_0 is given.

The total cost function can be written as

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

Recall that the focus in this paper is to evaluate the volume dependent shipment costs, $TSC(R_0, \mathbf{T}, \mathbf{w})$, and the emissions in the system. The inventory holding and backorder 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 reserved intermodal transportation capacity for each scheduled shipment. Apart from the actual cost of reserving capacity, this cost may include fixed costs for picking, receiving, administration, etc., (ii) fixed costs for the alternative transportation mode for each transportation unit/load carrier used, and (iii) linear costs per unit shipped for both transportation modes. 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 intermodal transportation mode

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

α_k'' = fixed cost for a single load carrier on the alternative transportation mode to retailer group k

c_i' = cost per unit shipped by the intermodal transportation option to retailer i

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

Δc_i = cost increase per unit for shipping with the alternative transportation mode to retailer i =
 $c_i'' - c_i'$

As each unit must be shipped on one of the two transportation modes, only the cost increase per unit for shipping with the alternative transportation mode, Δc_i , is needed 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 described and discussed in Appendix B.

It is worth emphasizing that 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 capacity of $w_k \in W_k$ is reserved on the intermodal transportation mode

β_k'' = fixed emissions for each load carrier or vehicle used on the alternative transportation mode to retailer group k

e_i' = emissions per unit shipped by the intermodal transportation mode to retailer i

e_i'' = emissions per unit shipped by the alternative transportation mode to retailer i

Δe_i = emissions increase per unit for shipping with the alternative transportation mode to retailer $i = e_i'' - e_i'$

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

The fixed emissions $\beta_k'(w_k)$ for the intermodal shuttle train option, are incurred regardless of how much of the reserved capacity that is actually used by the company. The motivation for this is that a shuttle train will cause emissions when it runs 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, we use the emissions increase per unit for shipping with the alternative transportation mode, Δe_i . We also assume that $\Delta e_{(k)} = \Delta e_i \forall i \in \Omega_k$ and discuss the relaxation of this assumption in Appendix B.

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 ($k=1, \dots, K$). 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,

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

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 $[t_0 - T_k, t_0)$,

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

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 $\lambda_0(t_0 - t_1)$.

$\text{mod}_{(R_0, Q_0)}(x) = x + nQ_0$, 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

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

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

$$IP_0(t) = \text{mod}_{(R_0, Q_0)}(IP_0(t_1) - D_0(t_1, t)) \quad (4)$$

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):

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

$t_1 - L_0 =$ the last point in time when orders placed by the warehouse will arrive in 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.

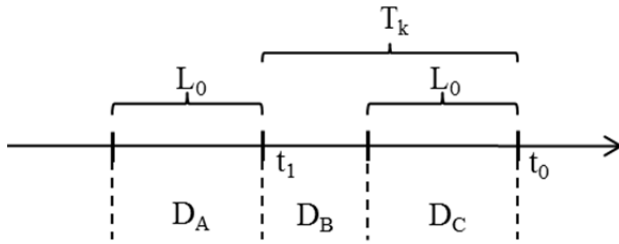


Figure 1. Time intervals when $L_0 \leq T_k$.

We start the analysis from $t_1 - L_0$, with an inventory position of $IP_0(t_1 - L_0) = x$. Note that this inventory position is uniformly distributed on $[R_0+1, R_0+Q_0]$ (see, for example, 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

$$IL_0(t_1) = x - D_A. \quad (5)$$

Similarly, the inventory level at time t_0 is

$$IL_0(t_0) = IP_0(t_0 - L_0) - D_C \quad (6)$$

From Lemma 2 we know that the inventory position at time $t_0 - L_0$ can be obtained from the inventory position x at time $t_1 - L_0$.

$$IP_0(t_0 - L_0) = \underset{(R_0, Q_0)}{\text{mod}}(x - D_A - D_B) \quad (7)$$

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

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

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). 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 $D_F = D_0(t_1, t_0)$.

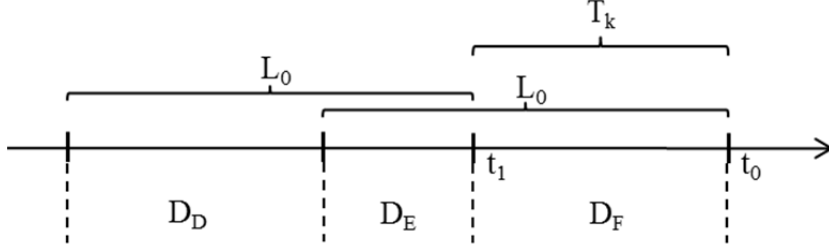


Figure 2. Time intervals when $L_0 > T_k$.

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

$$IL_0(t_1) = x - D_D - D_E. \quad (9)$$

From Lemma 2, we know that $IP_0(t_0 - L_0) = \text{mod}_{(R_0, Q_0)}(x - D_D)$, which implies that the inventory level at time t_0 is

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

M can, from Lemma 1, be obtained as

$$M = D_F + IL_0^-(t_1) - IL_0^-(t_0). \quad (11)$$

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 intermodal capacity, w_k , and the shipment quantity $m_{(k)}$

Based on the analysis of the shipment quantities in the previous section the expected shipment cost per time unit to retailer group k , TSC_k , can be obtained as

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

The total expected shipment cost per time unit follows directly

$$\text{TSC}(\mathbf{R}_0, \mathbf{T}, \mathbf{w}) = \sum_{k=1}^K \text{TSC}_k. \quad (13)$$

Given w_k and $m_{(k)}$, the number of units shipped with the alternative transportation mode is

$(m_{(k)} - w_k)^+$. The shipment cost is thus

$$\text{SC}_k(w_k, m_{(k)}) = \alpha_k'(w_k) + n_k \alpha_k'' + (m_{(k)} - w_k)^+ \Delta c_{(k)}, \quad (14)$$

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

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

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

where n_k is defined as in (14). The expected emissions per time unit for shipments to retailer group k , TE_k , is now obtained as

$$\text{TE}_k = \frac{1}{T_k} \sum_{m_{(k)}=0}^{\infty} P(M_{(k)} = m_{(k)}) E_k(w_k, m_{(k)}). \quad (16)$$

The total expected emissions per time unit follows

$$\text{TE}(\mathbf{R}_0, \mathbf{T}, \mathbf{w}) = \sum_{k=1}^K \text{TE}_k. \quad (17)$$

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 $\mathbf{R}_0, \mathbf{S}, \mathbf{T}, \mathbf{w}$ 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 maximum expected emissions per time unit allowed in the system, θ . The presented optimization method focuses on scenario (iii), as (i) and (ii) are special cases where θ is infinite, and for (ii) the original cost parameters are also modified to 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 $\{S_1, S_2, \dots, S_N\}$. To see this, note that the emissions and transportation costs associated with \mathbf{w} are unaffected by the order-up-to levels at the retailers when R_0 and \mathbf{T} are fixed. The result then follows directly from Marklund (2011).

The proposed optimization procedure can be described as a bounded search over all relevant combinations of \mathbf{T} , R_0 and \mathbf{w} ($w_k \in W_k \forall k=1, \dots, K$), using the convexity for optimizing \mathbf{S} (given \mathbf{T} and R_0). More precisely, for each relevant combination of \mathbf{T} , \mathbf{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 $\mathbf{S} = \{S_1, S_2, \dots, S_N\}$. 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, 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 $k=1, 2, \dots, K$), 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, $w_k \in W_k \forall k=1, \dots, K$.

It is clear that a critical step in the optimization procedure is to determine the bounds on the shipment intervals and warehouse reorder point. Starting with the latter, we can 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 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 bound ensures that an increase in R_0 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 sizes 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 noteworthy that it is not difficult to find tighter upper bounds on R_0 (see for example Stenius et al. 2012). However, because the fast recursive procedure in Marklund (2011) is initiated at R_0^u above, and in the process of evaluating the costs associated with R_0^l , it renders the cost optimal solution for all intermediate R_0 values, a tighter upper bound does not improve the computational performance.

Turning to the shipment intervals, an obvious lower bound is $T_k^l = T_{\min}$, $\forall k=1, \dots, K$. An upper bound for T_k is provided in Proposition 2.

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

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

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

It is noteworthy that Proposition 2 provides upper bounds on T_k for all k , provided that a feasible solution has been found and thus $\overline{TC} < \infty$ (note that in scenario (i) and (ii) all solutions are feasible). However, if θ is set too low, solutions fulfilling the emissions constraint may not exist. In order to determine in which cases feasible solutions exist (and to find a feasible solution) it might be necessary to find the lowest possible expected emissions per time unit, TE° . If $TE^\circ > \theta$, no feasible solution exists. If $TE^\circ \leq \theta$, a feasible solution can be found by optimizing S (using the aforementioned convexity property) for the system parameters R_0 , T and w corresponding to TE° . For reasons of exposition a further analysis of this special case and how to determine TE° 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 objective is to demonstrate how the presented model can be used for optimizing the system performance in terms of minimizing expected costs while meeting an emissions target (scenario (iii)). The example illustrates how the model can be used to understand the cost impact of managerial decisions regarding emissions targets. The cost parameters are motivated by discussions with industry and the emissions calculations are based on information from the NTM database (NTM, 2011).

The inventory system consists of a central warehouse (in central Europe) supplying three retailers in two retailer groups ($N=3$, $K=2$). One retailer group (in Scandinavia) consists of two retailers ($\Omega_1 = [1,2]$), and a single retailer (in southern Europe) constitutes the second retailer group ($\Omega_2 = [3]$). To both groups there exist opportunities to reserve an integer number of load carriers on an intermodal shuttle train solution with possible reservation quantities in number of item units, $W_1 = W_2 = [0,5,10,15,20]$. Thus, for each retailer group it is possible not to use the train option (reserve 0 units), or to reserve capacity to ship 5, 10, 15 or at most 20 units by train (corresponding to 1, 2, 3 or 4 load carriers containing 5 units each). The alternative for both retailer groups is to use direct truck transports that can be acquired by the load carrier of size $A_1 = A_2 = 5$ units. The associated fixed costs for the intermodal train solution is 1500 € per load carrier, which means that $\alpha_1'(W_1) = \alpha_2'(W_2) = [0,1500,3000,4500,6000]$ €. Thus it costs 3000 € to reserve capacity to ship 10 units by train, (even if fewer units are actually shipped). For the trucking option the cost per load carrier bought is 2000 €, i.e., $\alpha_1'' = \alpha_2'' = 2000$ €. There are also variable transportation costs per unit shipped that are 20 € higher for the trucking alternative ($\Delta c_{(1)} = \Delta c_{(2)} = 20$ €).

The emissions associated with the different transportation options are also divided into fixed and variable components. Fixed emissions for the different capacity reservations on the

train are $\beta_1'(W_1) = \beta_2'(W_2) = [0,200,400,600,800]$ kg CO₂ equivalents. This means that transporting an empty load carrier by train corresponds to emissions of 200 kg CO₂ equivalents. For the trucking alternative, the fixed emissions per load carrier are 850 kg CO₂ equivalents, i.e., $\beta_1'' = \beta_2'' = 850$. The variable emissions per unit shipped are 10 kg CO₂ equivalents higher for the trucking alternative than for the intermodal train option ($\Delta e_{(1)} = \Delta e_{(2)} = 10$).

The transportation time from the warehouse to each retailer is 2 days ($L_i = 2$), and the leadtime for warehouse orders is 10 days ($L_0 = 10$). The batch size for warehouse orders is 10 units ($Q_0 = 10$). Holding costs are the same at the central warehouse and at the retailers ($h_0 = h_i = 10$ €/day, for all i), and so are the backorder costs at the retailers ($b_i = 100$ €/day, for all i). The mean demand per day at each retailer is 0.5 units and the smallest shipment interval considered is 1 day ($T_{\min} = 1$ day). Note that the model is in no way restricted to identical parameters across retailers, it is used in this example only for expositional reasons.

Table 1 illustrates the results for three problem instances; (1) no emissions constraint exists, (2) The emissions target $\theta = 100$ kg CO₂ equivalents, and (3) $\theta = TE^c$. 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 CO₂ equivalents per day (from (17)). Enforcing the emission constraint of $\theta = 100$ kg CO₂ 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 1 shows that more capacity is now reserved on the intermodal transportation modes (\mathbf{w} increases from [10,5] to [15,10]), the shipment intervals (\mathbf{T}) are longer (to ensure that the reserved capacity is better utilized), and the order-up-to levels at the retailers (\mathbf{S}) are increased (to balance holding and backorder

costs for longer replenishment times). The system with the lowest possible expected emissions ($\theta = TE^c$), reserves maximum capacity on the intermodal transportation modes, and have 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 (TSC decreases further but TIC increases more).

Table 1. Optimal solutions for different θ values.

θ	R_0	T	w	S	TIC*	TSC*	TC*(θ)	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^c	14	[16,32]	[20,20]	[10,10,17]	447.09	588.95	1036.04	86.29

Table 2. Analysis of the resulting shipment solutions for different θ values.

θ	Ret grp	Shipment volume		Utilization of w_1, w_2	Probability of using x trucks				
		train	truck		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^c	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

Table 2 provides further analysis of shipment solutions for the three problem instances. Without an emissions constraint, train 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 $\theta = TE^c$, the train freight ratio increases to 98% leaving only 2% to be transported by truck. A closer look at the trucking 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 $\theta = TE^c$. (These probabilities are obtained directly from the pmf of the shipment size $M_{(k)}$ 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% ($\theta = TE^c$) as the emissions constraint is tightened. Intuition suggests that this is a result

of hedging more against having to use the trucking option. However, for retailer group 2, the capacity utilization first increases from 77% to 79% ($\theta=100$) and then decreases to 78% ($\theta=TE^c$). Consequently, the behavior is difficult to predict in a general situation. One reason is of course the integrality of the decision variables.

Figure 3 provides the lowest expected costs, $TC^*(\theta)$, for all θ values between TE^c 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 larger emissions cuts, the marginal cost increase escalates. It follows that the model can serve as a tool for evaluating the cost impact of managerial decisions to reduce allowable emissions, and for analyzing the tradeoff between reduced expected emissions and costs in the system. The model also prescribes how to achieve the emission targets at minimum cost by choosing optimal values of the decision variables, i.e., the length of the shipment intervals, $\mathbf{T}=\{T_1, \dots, T_K\}$, the warehouse reorder point, R_0 , the retailer base-stock levels, $\mathbf{S}=\{S_1, \dots, S_N\}$, and the capacity reservations per retailer group on the intermodal train solution, $\mathbf{w} = \{w_1, w_2\}$.

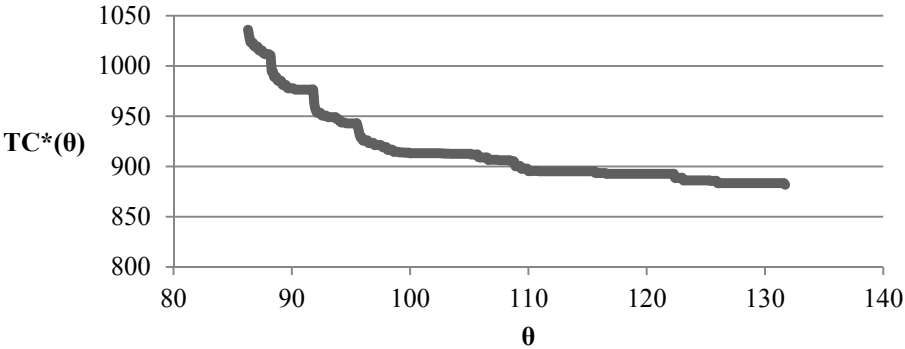


Figure 3. Expected Total Cost for varying θ values.

6. Extensions and generalizations

In this section we explain how the single-item model can be extended to multi-item systems. We also discuss generalizations to other cost and emissions structures.

6.1 Extension to multi-item systems

Clearly, in practice, shipment consolidation usually involves many different items that are shipped together. Thus the relevance of considering multi-item systems is quite obvious, particularly for low demand items. Fortunately, it is straightforward to extend the single-item model to multi-item systems. 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 cost and total emissions across all items and retailers in the system. An added complication is that transportation capacity is ambiguous to measure in number of units. Instead a common measure like weight or volume needs to be used when different items are shipped together.

To elaborate, 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 other retailers and of 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, $\mathbf{R}_0 = \{R_0^1, R_0^2, \dots, R_0^J\}$

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

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 intermodal transportation option,
- \mathbf{w} = vector of capacity reservation per shipment to each of the K retailer groups for the intermodal transportation option $\{w_1, w_2, \dots, w_K\}$, $w_k \in W_k \forall k=1, 2, \dots, K$
- h_i^j = holding cost per unit and time unit of item j at stock point i , $i = 0, 1, \dots, N$
- b_i^j = backorder cost per unit and time unit of item j at retailer i , $i = 1, 2, \dots, N$
- 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
- $TIC^j(\mathbf{R}_0, \mathbf{S}^j, \mathbf{T}, \mathbf{w})$ = 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

$$TC(\mathbf{R}_0, \tilde{\mathbf{S}}, \mathbf{T}, \mathbf{w}) = \sum_{j=1}^J TIC^j(\mathbf{R}_0, \mathbf{S}^j, \mathbf{T}, \mathbf{w}) + \sum_{k=1}^K TSC_k(\mathbf{R}_0, \mathbf{T}, \mathbf{w}) \quad (19)$$

For reasons of exposition, we assume that exactly w_k size units can be shipped with the intermodal transportation mode to retailer group k whenever $V_{(k)} \geq w_k$. As in Section 2, we also assume that $\Delta c_{(k)} = \Delta c_i$ and $\Delta e_{(k)} = \Delta e_i \quad \forall i \in \Omega_k$.

The probability mass function of the shipment quantities, $M_{(k)}^j$, can be analyzed independently for all units $j=1,2,\dots,J$ 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 then be obtained as

$$TSC_k = \frac{1}{T_k} \sum_{v_{(k)}=0}^{\infty} P(V_{(k)} = v_{(k)}) SC_k(w_k, v_{(k)}), \quad (20)$$

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 intermodal transportation mode. $SC_k(w_k, v_{(k)})$ is obtained as

$$SC_k(w_k, v_{(k)}) = \alpha_k'(w_k) + n_k \alpha_k'' + (v_{(k)} - w_k)^+ \Delta c_{(k)}, \quad (21)$$

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

$$E_k(w_k, v_{(k)}) = \beta_k'(w_k) + n_k \beta_k'' + (v_{(k)} - w_k)^+ \Delta e_{(k)}. \quad (22)$$

This gives expected emissions per time unit for shipments to retailer group k of

$$TE_k = \frac{1}{T_k} \sum_{v_{(k)}=0}^{\infty} P(V_{(k)} = v_{(k)}) E_k(w_k, v_{(k)}) \quad (23)$$

and total expected emissions per time unit of

$$TE(\mathbf{R}_0, \mathbf{T}, \mathbf{w}) = \sum_{k=1}^K TE_k. \quad (24)$$

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 $\tilde{\mathbf{S}} = \{S_1^1, S_2^1, \dots, S_N^1; S_1^2, S_2^2, \dots, S_N^2; \dots; S_1^J, S_2^J, \dots, S_N^J\}$. 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, given the emissions constraint, $TC^*(\theta)$, has been found and the minimum total expected 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 $j = 1, \dots, J$. 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 (Appendix C) and TE^e is obtained as

$$TE^e = \min_{\mathbf{r} \in \mathbf{R}^j} \sum_{k=1}^K TE_k^e(\mathbf{r}), \quad (25)$$

6.2 Other extensions

Thus far, we have assumed that the relative cost increase and the relative emissions per unit (or per size unit in the multi-item case) for using the alternative transportation mode is the

same for all retailers in a group, i.e. $\Delta c_i = \Delta c_{(k)}$ and $\Delta e_i = \Delta e_{(k)}$ for all $i \in \Omega_k$. If this assumption is relaxed and the shipment size exceeds the reserved capacity, there is a need to allocate units to different transportation modes based on their destination. In order to evaluate the effects of this 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 analysis and the associated allocation rules are described in Appendix B. The multi-item case can be handled analogously. Note that the joint pmf of the shipment quantities (see Appendix B), also enables evaluation of other systems with more complex transportation cost and emissions structures.

Relaxing the assumption made in the multi-item case, that exactly w_k size units can be shipped to retailer group k whenever $V(k) \geq w_k$ requires a more sophisticated allocation decision, which considers how the intermodal transportation mode is best utilized. Analyzing optimal allocation decisions in this setting is an interesting direction for future research.

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. 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 function for the size of an arbitrary shipment to each retailer group. With this distribution at hand, size-dependent transportation cost structures and emissions from transportations can be evaluated. Combined with the inventory holding and backorder costs at the warehouse and retailers, determined in Marklund (2011), a joint optimization of inventory replenishment and transportation decisions can be performed with respect to both costs and emissions. Single-item as well as multi-item settings can be analyzed and optimized.

To emphasize the sustainability aspects of our results the model encompasses opportunities to reserve transportation capacity on an intermodal shuttle train solution (truck-train-truck) from the central warehouse to at least one of the retailer groups. Units exceeding the reserved capacity are shipped with an alternative transportation mode (truck). Apart from derivations of exact probability mass functions of the shipment sizes, the contributions include determination of the exact expected costs and emissions. Also, procedures for joint optimization of reorder points, shipment intervals, and capacity reservation on the intermodal train solution are presented. A numerical example illustrates how cost optimization subject to an emissions constraint can be performed. It is shown how the model can be used for evaluating the cost impact of managerial decisions to impose emissions targets.

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. Also, it would be interesting to evaluate other consolidation policies.

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 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 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.
- Benjaafar, S., Y. Li, M. Daskin. 2013. Carbon footprint and the management of supply chains: Insights from simple models. *IEEE Transactions on Automation Science and Engineering* **10** 99-116.

- Berling P., F. Eng-Larsson. 2013. Reserving capacity to enable green investment - effects on technology choice, inventory, and environmental performance. Working paper, Lund University.
- Bouchery, Y., A. Ghaffari, Z. Jemai, Y. Dallery. 2012. Including sustainability criteria into inventory models. *European Journal of Operational Research* **222** 229-240.
- 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.P., M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science* **46** 1032-1048.
- Caro, F., C.J. Corbett, T. Tan, R. Zuidwijk. 2013. Double counting in supply chain carbon footprinting. *Manufacturing & Service Operations Management* **15** (4) 545-558.
- 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., F. Mutlu, C.-Y. Lee. 2006. A comparison of outbound dispatch policies for integrated inventory and transportation decisions. *European Journal of Operational Research* **171** 1094-1112.
- 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 Operations Research* **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.
- Corbett, C.J., P.R. Kleindorfer. 2001a. Environmental management and operations management: introduction to part 1 (manufacturing and eco-logistics). *Production and Operations Management* **10** 107-111.
- Corbett, C.J., P.R. Kleindorfer. 2001b. Environmental management and operations management: introduction to part 2 (integrating operations and environmental management systems). *Production and Operations Management* **10** 225-227.
- 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.
- Dogru, M.K. 2006. Optimal control of one-warehouse multi-retailer systems: An assessment of the balance assumption. Doctoral Thesis, Beta Research School for Operations Management and Logistics, Technische Universiteit Eindhoven.
- Federgruen, A., P. Zipkin. 1984a. Approximations of dynamic multilocation production and inventory problems. *Management Science* **30** 69-84.
- Federgruen, A., P. Zipkin. 1984b. Allocation policies and cost approximations for multilocation inventory systems. *Naval Research Logistics Quarterly* **31** 97-129.
- 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. Moınzadeh, 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.
- Hoen, K.M.R., T. Tan, J.C. Fransoo, G.J. van Houtum. 2012. Effect of carbon emission regulations on transport mode selection under stochastic demand. *Flexible Services and Manufacturing Journal*. DOI 10.1007/s10696-012-9151-6
- Hoen, K.M.R., T. Tan, J.C. Fransoo, G.J. van Houtum. 2013. Switching transport modes to meet voluntary carbon emission targets. Working paper Eindhoven University of Technology.
- 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.
- Hua, G., T.C.E. Cheng, S. Wang. 2011. Managing carbon footprints in inventory management. *International Journal of Production Economics* **132** 178–185.
- Jaber, M.Y., C.H. Glock, A.M.A. El Saadany. 2013. Supply chain coordination with emission reduction incentives. *International Journal of Production Research* **51** 69-82.
- 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., K. Rosling. 2012. Lower bounds and heuristics for supply chain stock allocation. *Operations Research* **60** 92–105.
- Moınzadeh K. 2002. A multi-echelon inventory system with information exchange. *Management Science* **48** 414–426.
- Mutlu, F., S. Çetinkaya. 2010. An integrated model for stock replenishment and shipment scheduling under common carrier dispatch costs. *Transportation Research Part E* **46** 844–854.
- 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.H., S.X. Zhou. 2012. Optimizing replenishment intervals for two-echelon distribution systems with stochastic demand. Working paper, Fuqua School of Business, Duke University, North Carolina, USA.
- 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. 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.
- Toth, P., and D., Vigos, *The Vehicle Routing Problem*, SIAM, Philadelphia, 2001.
- Verrijdt J.H.C.M., A.G. de Kok, 1996. Distribution planning for divergent depotless two-echelon network under service constraints. *European Journal of Operational Research* **89** 341–354.

Appendix A - Proofs of propositions and lemmas

Proof of Proposition 1. Because the retailer order processes are independent Poisson processes, and because of the FCFS allocation at the central warehouse, the probability that an arbitrary unit that becomes 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 size $m_{(k)}$ to retailer group k given m qualified units in total, will be $\text{Bin}(m_{(k)}, m, \lambda_{(k)}/\lambda_0)$. Taking the expectation of all possible values of M renders (2).

Proof of Lemma 1. Because all demanded units eventually will be qualified for shipment, we know that 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 - Other cost and emissions structures, and the joint pmf of shipment quantities

Whenever the shipment quantity exceeds the reserved capacity to retailer group k , an allocation decision arises concerning which units (to which retailers within retailer group k) should be shipped using the intermodal transportation option. This decision is avoided in the main analysis as it is assumed that the cost increase, Δc_i , and the increase in environmental impact, Δe_i , are the same for all retailers within a retailer group. To perform the allocation, the joint probability mass function of the shipment quantities to all retailers within retailer group k is needed. For reason of exposition and without loss of generality, we renumber the retailers in retailer group k $\{1, 2, \dots, N_k\}$, and renumber the rest of the retailers from $N_k + 1$ to N . We define:

$M_i(t_0) =$ 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), $M_i \equiv M_i(t_0)$

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

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

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

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 general 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 , it follows that

$$P\{\mathbf{M}_k = \mathbf{m}_k | M = m\} = \frac{m!}{(m - \varphi)! \prod_{i=1}^{N_k} m_i!} \left(\frac{\lambda_0 - \lambda_{(k)}}{\lambda_0} \right)^{(m - \varphi)} \prod_{i=1}^{N_k} \left(\frac{\lambda_i}{\lambda_0} \right)^{m_i}, \quad (\text{B1})$$

and

$$P\{\mathbf{M}_k = \mathbf{m}_k\} = \sum_{m=\varphi}^{\infty} P\{M = m\} P\{\mathbf{M}_k = \mathbf{m}_k | M = m\}, \quad (\text{B2})$$

With the joint probability mass function of the shipment quantities to all retailers in retailer group k known, we turn our attention to the allocation decision. The allocation of units on different transportation modes is dependent on the optimization objective. For a cost optimization without environmental considerations (scenario (i) in Section 4) the allocation is straightforward: the intermodal transportation mode should be prioritized for the retailers with the highest cost increase for shipping with the alternative transportation mode, Δc_i . For systems where there is a known internal cost for the environmental impact imposed, ρ , (scenario (ii) in Section 4) the allocation rule is similar; the capacity on the intermodal transportation mode should be allocated to the retailers with the largest marginal cost $\Delta c'_i = \Delta c_i + \rho \cdot \Delta e_i$. For scenario (iii), where there is a hard constraint on the expected emissions per time unit for the system, the optimal allocation rules can become complex. Situations might occur where the same allocation decision (decisions for the same shipment quantities to all retailers) should be performed differently at different time instances. These and other complicating features are left for future research. We instead focus the analysis on

the shipment costs and environmental impact in scenario (ii). Scenario (i) can be seen as the special case of this scenario where $\rho = 0$.

Based on the analysis of the shipment size distributions, the expected shipment cost per time unit to retailer group k can then be obtained as

$$\text{TSC}_k = \frac{1}{T_k} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \dots \sum_{m_{N_k}=0}^{\infty} P(\mathbf{M}_k = \mathbf{m}_k) \text{SC}_k(w_k, \mathbf{m}_k). \quad (\text{B2})$$

Without loss of generality, we renumber the retailers in retailer group k so that $\Delta c'_1 \geq \Delta c'_2 \geq \dots \geq \Delta c'_{N_k}$, and obtain the shipment cost for a shipment, given w_k and $\mathbf{M}_k = \mathbf{m}_k$, as

$$\text{SC}_k(w_k, \mathbf{m}_k) = \alpha_k'(w_k) + \begin{cases} 0 & \text{if } \sum_{i=1}^{N_k} m_i \leq w_k \\ n\alpha_k'' + \left(\sum_{i=1}^j m_i - w_k \right) \Delta c_j + \sum_{i=j+1}^{N_k} m_i \Delta c_i & \text{if } w_k + (n-1)A_k < \sum_{i=1}^{N_k} m_i \leq w_k + nA_k, \end{cases} \quad (\text{B3})$$

where $n = 1, 2, \dots$ and j is decided so that

$$\sum_{i=1}^{j-1} m_i \leq w_k < \sum_{i=1}^j m_i. \quad (\text{B4})$$

Analogously, we obtain the expected emissions for a shipment, given w_k and $\mathbf{M}_k = \mathbf{m}_k$, as

$$E_k(w_k, \mathbf{m}_k) = \beta_k'(w_k) + \begin{cases} 0 & \text{if } \sum_{i=1}^{N_k} m_i \leq w_k \\ n\beta_k'' + \left(\sum_{i=1}^j m_i - w_k \right) \Delta e_j + \sum_{i=j+1}^{N_k} m_i \Delta e_i & \text{if } w_k + (n-1)A_k < \sum_{i=1}^{N_k} m_i \leq w_k + nA_k. \end{cases} \quad (\text{B5})$$

This gives the expected environmental impact per time unit for shipments to retailer group k ,

TE_k , as

$$\text{TE}_k = \frac{1}{T_k} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \dots \sum_{m_{N_k}=0}^{\infty} P(\mathbf{M}_k = \mathbf{m}_k) E_k(w_k, \mathbf{m}_k). \quad (\text{B6})$$

Optimization may be carried out using the same approach as described in Section 4 (for scenarios (i) and (ii)).

Appendix C - 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

$$TE^e = \min_{r=R_0^l, \dots, R_0^u} \sum_{k=1}^K TE_k^e(r). \quad (C1)$$

The lowest emissions in retailer group k given $R_0=r$, $TE_k^e(r)$, are found by searching through T_k , increasing it incrementally from T_k^l . 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 C1 below, we stop increasing T_k when TE_k gets sufficiently close to $\lambda_{(k)} \left(\Delta e_{(k)} + \beta_k'' / A_k \right)$ for all $w_k \in W_k$.

Lemma C1.

$$\lim_{T_k \rightarrow \infty} TE_k = \lambda_{(k)} \left(\Delta e_{(k)} + \beta_k'' / A_k \right). \quad (C2)$$

Proof. Whenever $(m_{(k)} - w_k)^+$ is not a multiple of A_k (where $m_{(k)}$ is the shipment size), 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,

$$\begin{aligned}
\lim_{T_k \rightarrow \infty} TE_k &= \lim_{T_k \rightarrow \infty} \frac{\beta'_k(w_k) + \sum_{m_{(k)}=0}^{\infty} P(M_{(k)} = m_{(k)}) \left[(m_{(k)} - w_k)^+ \Delta e_{(k)} + n_k \beta''_k \right]}{T_k} \\
&= \lim_{T_k \rightarrow \infty} \frac{\sum_{m_{(k)}=0}^{\infty} P(M_{(k)} = m_{(k)}) \left[(m_{(k)} - w_k) \Delta e_{(k)} + \frac{(m_{(k)} - w_k)}{A_k} \beta''_k + \frac{Y}{A_k} \beta''_k \right]}{T_k} \\
&= \lim_{T_k \rightarrow \infty} \frac{E \left[m_{(k)} - w_k \right] \left(\Delta e_{(k)} + \frac{\beta''_k}{A_k} \right) + E[Y] \frac{\beta''_k}{A_k}}{T_k} \\
&= \lim_{T_k \rightarrow \infty} \frac{(\lambda_{(k)} T_k - w_k) \left(\Delta e_{(k)} + \frac{\beta''_k}{A_k} \right)}{T_k} = \lambda_{(k)} \left(\Delta e_{(k)} + \frac{\beta''_k}{A_k} \right).
\end{aligned} \tag{C3}$$

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 size smaller than w_k approaches zero as T_k approaches infinity. The fourth and fifth equalities follow as both

$$E[Y] \frac{\beta''_k}{A_k} \text{ and } -w_k \left(\Delta e_{(k)} + \frac{\beta''_k}{A_k} \right) \text{ are finite for } w_k \in W_k.$$

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

May 8, 2012

Olof Stenius

Department of Industrial Management and Logistics, Lund University, Lund, Sweden, Olle.Stenius@iml.lth.se

Gönül Karaarslan

Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands, A.G.Karaarslan@tue.nl

Johan Marklund

Department of Industrial Management and Logistics, Lund University, Lund, Sweden, Johan.Marklund@iml.lth.se

A.G. de Kok

Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands, 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 to groups of retailers periodically. Customer demand is compound Poisson distributed and unsatisfied demand at each stock point 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, Q) 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 batch ordering. 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 rising 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 consisting of one central warehouse and multiple non-identical retailers facing compound Poisson demand. The system is characterized by a time-based dispatching and shipment consolidation policy at the warehouse, in conjunction with real-time point of sale data and centralized inventory information. In our model this means that inventories are reviewed continuously, and base-stock policies are used to immediately convey demand information from the different retailers to the warehouse. The warehouse on the other hand uses a (R, Q) policy (a batch of Q units is ordered when the warehouse inventory position falls to or below the reorder point R) to replenish its stock from an outside supplier/manufacturer. Batch ordering at the central warehouse is motivated by set up costs and batch restrictions at the manufacturer. Shipments from the central warehouse are consolidated for groups of retailers and dispatched periodically. The time between shipments to a retailer group is referred to as a shipment interval. These intervals may be different for different retailer groups and items. For each shipment there is a fixed shipment cost. There are also linear holding costs at all inventory locations together with either fill rate constraints or backorder costs per unit and time unit at the retailers. All locations employ complete backordering and First-Come First-Served (FCFS) allocation policies.

We find exact expressions for the probability distributions of inventory levels, from which we derive the expected costs for this distribution system. Using properties of the expected costs, we present

methods to jointly optimize shipment intervals, reorder points, and base-stock levels to reach a cost-optimal solution. As indicated above, we consider both the case of backorder costs per unit and time unit at the retailers (i.e., the total cost encompass expected shipment, holding and backorder costs) and the case of fill rate constraints at the retailers (where total cost consists of the expected shipment and holding costs only). The presented approach is applicable both for single-item and multi-item systems, although the analysis presented in this paper focuses on the former.

The motivation for studying the described model stems from discussions and interaction with several industrial companies that use different types of VMI (Vendor Managed Inventory) systems (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 (here modeled as a central warehouse) with the mandate to control inventories at different customer locations (here modeled as 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 enable the supplier to be flexible in planning the production, distribution and replenishing activities. Another motivation for our work is that the literature is scarce on general multi-echelon inventory models that explicitly incorporate shipment consolidation decisions. Thus there is a need for such models, not least as tools for achieving more sustainable inventory distribution 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 points at all inventory locations for a given set of shipment intervals. It also presents a heuristic for determining near optimal shipment intervals. 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 us to compute inventory level distributions and service levels at all locations, not just the expected costs as Marklund (2011). Thirdly, we provide upper and lower bounds for the joint optimization of shipment intervals and reorder points at all locations in the system facilitating an exact optimization procedure. Howard and Marklund (2011) build on the model in Marklund (2011) and use simulation to investigate the

cost benefits of replacing the First-Come First-Served (FCFS) allocation assumption at the supplier with a state dependent myopic allocation policy. The conclusion is that some cost benefits exist for long lead times, particularly, when the allocation decision is postponed to the moment of delivery, but in general the FCFS assumption performs very well. This result supports the use of FCFS allocation in our model. Gürbüz et al. (2007) also consider joint inventory and transportation decisions, but for a system where the central warehouse is not allowed to carry inventory. More precisely, they consider a centralized ordering policy for a cross docking facility, which orders for a set of retailers simultaneously. This policy is compared to three well-known policies including a fixed replenishment interval policy.

Our 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 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 based on convolutions and binomial disaggregation of warehouse backorders to obtain the inventory level distributions at the retailers. Knowing these distributions it is straight forward to determine the expected holding and backorder costs per time unit for the system. Axsäter (1990) provides fast and exact recursive expressions for the expected inventory holding and backorder costs by tracking an arbitrary unit through the system. Our present work is related to both these approaches as we derive the exact distributions of warehouse backorders designated to the different retailers 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. Axsäter (1997) extends the model to echelon stock (R, Q) -policies (the echelon stock of an inventory location is defined by the installation stock [or local stock] of the location itself plus the installation stock levels of all downstream stages). In a parallel work Chen and Zheng (1997) provide an alternative method for evaluating echelon-stock (R, Q) -policies which 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 present 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 multiple-retailer system with installation stock (R, Q) -policies and compound Poisson demand. In the special case of no shipment consolidation (i.e. all shipment intervals = 0) and order quantities equal to one at all retailers, our present work offers an alternative way to analyze the system in Axsäter (2000). It is noteworthy that our method of analysis is very different from the approach used in Axsäter (2000).

The same research stream on exact analysis of divergent continuous review systems also contains a number of recent papers that investigate more complex warehouse ordering policies. These policies utilize more detailed inventory and demand information than the traditional base-stock or (R, Q) policies to improve the inventory control. 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). This replenishment policy uses real-time information about the individual inventory positions at all stock points. 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 present in all previous models cited above. Marklund (2006) focus on the use of advance order (or demand) information, and provides exact and approximate analysis of warehouse reservation policies. These policies enable the warehouse to differentiate its service across the retailers through temporal allocation and prioritization. All the models above assume the use of partial deliveries, i.e., all stock points 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 multiple retailer distribution systems, see, for example, Axsäter (2003) for an overview. One of the most common approximation techniques is to replace the stochastic waiting time due to stockouts at a stock point by its average value. This idea originates with 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, Axsäter and Marklund (1998), Andersson and Marklund (2000) and Axsäter (2003). 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. They derive an exact expression for the waiting time distribution of an arbitrary customer under compound renewal demand and installation (R, Q) -policy. From this exact, yet computationally intractable, expression they develop accurate approximations for the first two moments of the waiting time of replenishment orders at their supplying stock point. Assuming replenishment orders must be filled completely before being shipped, they find the first two moments of the so-called effective lead time being the sum of replenishment lead time and waiting time. The resulting approximations for N-echelon divergent systems under compound renewal demand and installation (R, Q) -policies are shown to be quite accurate.

The time based dispatching and use of fixed shipment intervals also 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 a,b), Jackson (1988), Federgruen (1993), Axsäter (1993b), Graves (1996), Verrijdt and de Kok (1996), Houtum, Inderfurth and Zijm (1996), Heijden, Diks and de Kok (1997), Diks and de Kok (1998,1999), Cachon (1999), Cachon and Fisher (2000), Axsäter, Marklund and Silver (2002), Özer (2003), Dogru (2006), Chu and Shen (2010), Shang and Zhou (2011), Marklund and Rosling (2012), and references therein) is that in our current model only shipments are made periodically, while inventory is reviewed and replenished continuously. Among the periodic review papers Graves (1996), Axsäter (1993b), and Shang and Zhou (2012) have a closer relationship with our work because they assume a virtual (FCFS) allocation policy based on Poisson demand arrivals. In order to use this policy, 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 and Zhou (2012) 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. Our approach for optimizing the shipment intervals, reorder points and base-stock levels, is also based on deriving optimality bounds.

Finally, there is a connection between our work and the growing 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. (2006), Çetinkaya et al. (2008), Mutlu and Çetinkaya (2010), and Mutlu, Çetinkaya and Bookbinder (2010). These papers essentially consider three different consolidation strategies for Poisson or compound Poisson demand: (i) time-based shipment consolidation policies where shipments leave with certain time intervals, (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 since the last dispatch has passed. None of these papers provide an exact analysis for time-based shipment consolidation and inventory control under compound Poisson demand similar to our present work.

The remainder of this paper is organized as follows. In Section 2, we describe 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 5 extends the model formulation, cost analysis and optimization procedure to the multi-item case. Section 6 presents numerical results, and finally, Section 7 concludes.

2 Problem Formulation Single-Item Model

As indicated above, we consider a centralized continuous review inventory system with one-warehouse and N non-identical retailers. Initially (Sections 2, 3 and 4), we limit our attention to single-item systems. Extensions to multi-item systems are described in Section 5. The retailers order solely from the central warehouse (no lateral transshipments between retailers are allowed) and the warehouse replenishes its inventory from an outside supplier. The customer demand occurs at the retailers and unsatisfied demand is backordered. Units are allocated to demand according to a FCFS principle and

partial deliveries of orders are allowed.

The retailers are grouped such that shipments to the retailers within the same group are consolidated, i.e. the same vehicle(s) are used to ship the goods to all retailers in the group. We assume that the retailer groups are predefined, and we have K retailer groups in total (by definition $K \leq N$). The set of retailers belonging to retailer group k is denoted by N_k . There are fixed shipment intervals for delivering units from the central warehouse to each retailer group. The shipment interval for retailer group k , T_k , is the time between two consecutive shipments and \mathbf{T} is the vector of all shipment intervals (T_1, \dots, T_K) . The shipment intervals may be different across retailer groups. The shipment time from the warehouse to retailer i , L_i , (including picking, loading, transporting and receiving) is assumed to be constant. However, 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. The replenishment lead time for the central warehouse, L_0 , defined as the time from placement of an order until it is available at the central warehouse, is constant.

We assume that the central warehouse has access to point of sale data, i.e. the customer demand at the retailers is directly transferred to the warehouse. Moreover, due to the FCFS allocation policy, the moment a retailer places an order (triggered by a customer demand of the same size) units are reserved to satisfy this particular order. When the reserved units are on hand at the central warehouse they will be shipped with the next shipment leaving for the retailer in question. As a consequence, the inventory at the central warehouse can conceptually be divided in two different parts; the *availability stock* and the *consolidation stock*. The *availability stock* consists of two parts; (i) *available* units that are on hand at the central warehouse which have not yet been reserved to satisfy a particular retailer order, and (ii) *backordered* units that are not yet on hand at the central warehouse but have been reserved for a particular retailer. A unit that is both reserved and available becomes qualified for shipment, and while waiting to be shipped it belongs to the *consolidation stock*. In other words, the *consolidation stock* consists of units that are physically at the central warehouse, reserved for a particular retailer and awaiting the next shipment leaving to the retailer in question. The distinction between the two different types of warehouse inventories is important for the subsequent analysis, but clearly there is no physical handling when a unit moves from the *availability stock* to the *consolidation stock*.

The inventory level of a retailer is defined as the stock on hand minus the amount of backorders at

this retailer. The inventory level of the *availability stock* is defined analogously as the stock on hand of available (unreserved) units minus the backorders at the central warehouse. For the *consolidation stock* the inventory level is just the stock on hand of reserved units at the central warehouse (as there can be no backorders at the *consolidation stock*). The inventory position at a retailer is defined as its inventory level plus all outstanding orders. The inventory position at the central warehouse is defined as the inventory level of the *availability stock* plus all outstanding orders. We use the following notation to express the system stock levels and demand structure:

$IL_i(t)$: Inventory level at retailer i at time t (i.e., stock on hand minus the amount of backorders at retailer i)

$IL_{av}(t)$: Inventory level of the *availability stock* at the warehouse at time t

$IL_{con}(t)$: Inventory level of the *consolidation stock* at the warehouse at time t

$IP_i(t)$: Inventory position of retailer i at time t

$IP_0(t)$: Inventory position of the warehouse at time t

$D_i(t_1, t_2)$: 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 $D_i \equiv D_i(1)$)

ρ_i : Variance-to-mean ratio of the demand per time unit at retailer i ($\rho_i = Var[D_i]/E[D_i]$)

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

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

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

The customer demand for a given item at each retailer is assumed to follow a stationary compound Poisson process with a discrete compounding distribution. The demand processes can be different across retailers, but are assumed to be mutually independent. More precisely, customers demanding a given item arrive to retailer i with Poisson arrival rate λ_i and each customer demands a stochastic number of units Y_i . We assume that a customer always demands a positive number of units and the expected demand size is $\mu_i = E[Y_i]$ for all $i \in N$.

We denote the aggregate order arrival rate at the warehouse by λ_0 , and the size of an arbitrary retailer order at the warehouse by Y_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 $i \in N$. The arrivals of retailer orders to the central warehouse from all other retailers than i has an arrival rate of $\lambda_{ic} = \sum_{j \in N \setminus i} \lambda_j$ and an order size of $Y_{ic} = \sum_{j \in N \setminus i} \chi_j Y_j$.

Based on the definitions above, the probability mass function (pmf) of the demand during x time units at retailer i is

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

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 . Y_0^n and Y_{ic}^n is defined and determined analogously.

Turning to the inventory analysis, we define IL_i^+ as the stock on hand and IL_i^- as the backorders at retailer i . The stock on hand at the central warehouse, denoted IL_0^+ , is the sum of the stock on hand in the *availability stock*, IL_{av}^+ , and in the *consolidation stock*, IL_{con} . Backorders at the central warehouse only exist in the *availability stock* and are denoted B_0 .

The central warehouse uses a (R_0, Q_0) policy to replenish its stock. This means that as soon as the inventory position falls to or below R_0 an order of Q_0 units is placed with an outside supplier/manufacturer. The order quantity Q_0 is presumed to be given with the restriction that it must be a positive integer (it may, for example, be determined by a deterministic EOQ method, as suggested in Zheng 1992 and Axsäter 1996). The analysis of the inventory levels and costs may be modified to other replenishment policies at the central warehouse with known lead time and inventory position distributions, for example periodic order-up-to S policies or (s,S) policies.

All retailers use base-stock ordering policies with order-up-to level S_i for retailer i , and \mathbf{S} representing the vector of order-up-to levels (S_1, \dots, S_N) . The use of base-stock policies is a consequence of the assumption that there are no fixed costs for placing orders and that the demand information is immediately transferred to the warehouse. Batching is done centrally by consolidating shipments.

We consider inventory holding and backorder costs per unit and time unit, and fixed shipment costs associated with every scheduled shipment. More precisely, h_0 and h_i denote the holding cost per unit and time unit of the central warehouse and retailer i respectively. Similarly, β_i denotes the backorder cost per unit and time unit at retailer i , and ω_k represents the shipment cost for each

scheduled shipment to retailer group k .

The objective is to minimize the long-run average system-wide costs per time unit by optimizing the reorder points, order-up-to-levels and shipment intervals. The total cost function is denoted by $TC(R_0, \mathbf{S}, \mathbf{T})$ and it consists of three parts: (i) The expected warehouse holding cost per time unit, $h_0E[IL_0^+]$, (ii) the expected holding and backorder costs per time unit at retailer i , $h_iE[IL_i^+]$ and $\beta_iE[IL_i^-]$, and (iii) the shipment cost per time unit $SC(\mathbf{T})$. Clearly, the shipment cost is independent of R_0 and \mathbf{S} and only depends on the shipment intervals. $TC(R_0, \mathbf{S}, \mathbf{T})$ can be expressed as

$$TC(R_0, \mathbf{S}, \mathbf{T}) = h_0E[IL_0^+] + \sum_{i \in N} (h_iE[IL_i^+] + \beta_iE[IL_i^-]) + \sum_{k \in K} \frac{\omega_k}{T_k}. \quad (2)$$

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 $\beta_iE[IL_i^-]$ disappear from (2). The fill rate is defined as the portion of the total demand that can be fulfilled immediately 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 , \mathbf{S} and \mathbf{T} . In Section 3.1 we determine the average stock on hand 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 probability mass functions. Finally, Section 3.4 provides an illustrative numerical example. All proofs are deferred to Appendix B.

3.1 The Stock on Hand at the Central Warehouse

As explained above, the stock on hand at the central warehouse, IL_0^+ , is the sum of the stock on hand in the *availability stock*, IL_{av}^+ , and the *consolidation stock*, IL_{con} :

$$E[IL_0^+] = E[IL_{av}^+] + E[IL_{con}]. \quad (3)$$

The *availability stock* can be analyzed as a single-echelon system without shipment consolidation as the consequences of the latter is dealt with in the *consolidation stock*. Thus, in steady state $IL_{av} = IP_0 - D_0(L_0)$. Moreover, because the retailers use continuous review base-stock policies the order process at the central warehouse is a superposition of the compound Poisson demand process at the retailers. Assuming that not all orders seen by the warehouse (i.e., 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). The average physical stock of *available* units at the central warehouse is therefore

$$E[IL_{av}^+] = \frac{1}{Q_0} \sum_{IP_0=R_0+1}^{R_0+Q_0} E[(IP_0 - D_0(L_0))^+]. \quad (4)$$

As soon as a unit is both demanded by a retailer and available at the central warehouse (i.e., the moment it would be shipped if there was no shipment consolidation), the unit is conceptually transferred from the *availability stock* to the *consolidation stock* where it waits for the next shipment to leave.

Proposition 1. *The expected stock on hand in the consolidation stock is obtained as*

$$E[IL_{con}] = \frac{1}{2} \sum_{k \in K} \sum_{i \in N_k} \lambda_i \mu_i T_k. \quad (5)$$

3.2 Inventory Levels at the Retailers

To derive the probability mass function for the inventory level at a given retailer, consider a shipment designated to retailer group k that leaves the warehouse at time t_0 . This shipment arrives at retailer i in group k at time $t_0 + L_i$ and the next shipment will arrive at $t_0 + L_i + T_k$. We call this time interval $(t_0 + L_i, t_0 + L_i + T_k]$ a replenishment cycle for retailer i in group k (see Figure 1). This cycle will continuously repeat itself as time progresses.

The inventory level for retailer i at $t_0 + L_i + t$ ($0 < t \leq T_k$) is determined by the inventory level at $t_0 + L_i$, just after the considered shipment has arrived, and the demand in $(t_0 + L_i, t_0 + L_i + t]$.

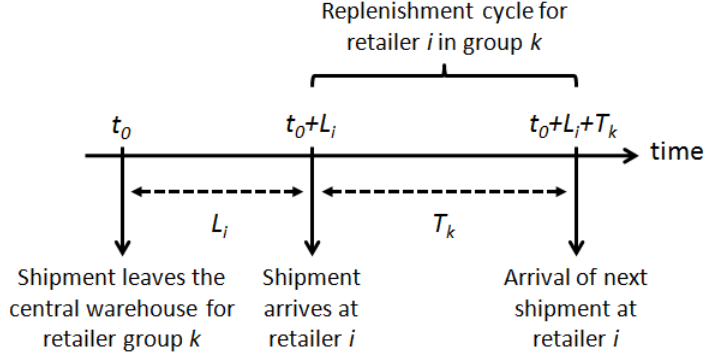


Figure 1: Replenishment cycle for retailer i in group k

Defining $B_i(t_0)$ as the number of backordered units allocated to retailer i at t_0 , we have

$$\begin{aligned}
 IL_i(t_0 + L_i + t) &= IL_i(t_0 + L_i) - D_i(t_0 + L_i, t_0 + L_i + t) \\
 &= S_i - B_i(t_0) - D_i(t_0, t_0 + L_i + t), \quad 0 < t \leq T_k.
 \end{aligned} \tag{6}$$

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 t_0 . By using this property, we derive the pmf of the inventory level of retailer i at $t_0 + L_i + t$ as

$$P\{IL_i(t_0 + L_i + t) = j\} = \sum_{r=0}^{S_i-j} P\{B_i(t_0) = r\} P\{D_i(L_i + t) = S_i - j - r\}, \quad j \leq S_i. \tag{7}$$

The challenging part in equation (7) is to determine the probabilities $P\{B_i(t_0) = r\}$. We focus on this in Section 3.3. For now we assume that these probabilities are known and concentrate on determining the expected stock on hand and the expected backorders at retailer i . To obtain the former, we determine the total expected stock on hand during a replenishment cycle and divide it by the total cycle length T_k , see (8).

$$\begin{aligned}
 E[IL_i^+] &= \frac{1}{T_k} \int_0^{T_k} E[IL_i(t_0 + L_i + x)^+] dx \\
 &= \frac{1}{T_k} \sum_{j=1}^{S_i} j \int_0^{T_k} P\{IL_i(t_0 + L_i + x) = j\} dx \\
 &= \frac{1}{T_k} \sum_{j=1}^{S_i} \sum_{r=0}^{S_i-j} j P\{B_i(t_0) = r\} \int_0^{T_k} P\{D_i(L_i + x) = S_i - j - r\} dx
 \end{aligned} \tag{8}$$

The expected inventory level can be determined analogously,

$$E[IL_i] = S_i - E[B_i(t_0)] - \lambda_i \mu_i \left(L_i + \frac{T_k}{2} \right). \quad (9)$$

The expected number of backorders at retailer i follows from (10).

$$E[IL_i^-] = E[IL_i^+] - E[IL_i] \quad (10)$$

The fill rate for retailer i , denoted by γ_i , can then be obtained from (11). 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.

$$\gamma_i = \frac{E[IL_i(t_0 + L_i)^+] - E[(IL_i(t_0 + L_i) - D(t_0 + L_i, t_0 + L_i + T_k))^+]}{\lambda_i \mu_i T_k} \quad (11)$$

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.

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 probability mass function (pmf) of $B_i(t_0)$, the number of backordered units designated to retailer i . As there can be no backorders in the *consolidation stock* (see Section 2), the analysis focus on the *availability stock*. 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

$$P\{B_i(t_0) = r\} = \frac{1}{Q_0} \sum_{S_0=R_0+1}^{R_0+Q_0} P\{B_i(t_0) = r | IP_0(t_0 - L_0) = S_0\} \quad (12)$$

The remaining analysis focus 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 $S_0 > 0$

We first note 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 . The total amount of units backordered at time t_0 , $B_0(t_0)$, will be

$$B_0(t_0) = (D_0(L_0) - S_0)^+. \quad (13)$$

Because of the FCFS policy, we also know that it always will 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 $(t_0 - L_0, t_0]$ (Recall that because of the base-stock policies at the retailers, every customer demand at a retailer will generate a retailer order to the warehouse of the same size). 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 ,

$$\Psi^n = S_0 - Y_0^n. \quad (14)$$

The nominal inventory position helps us to track which retailer orders are backordered at time t_0 . Note that $\Psi^0 = S_0$ and $\Psi^{\Phi_0} = IL_{av}(t_0)$, 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.

Focusing first on the partial backorders. We note that the distribution of the partial backorders belonging to retailer i will depend on which retailer order that causes these backorders. We define:

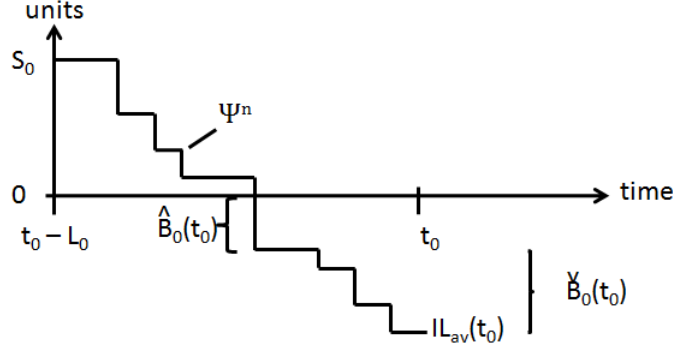


Figure 2: Illustration of a possible sample path for the nominal inventory position, Ψ^n , at the central warehouse and the associated number of partial and complete backorders

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

Lemma 2. *The probability of u partial backorders for retailer i at time t_0 , when the n^{th} customer order after $t_0 - L_0$ brings the nominal inventory position to 0 or below, $P\{\hat{B}_i^n(t_0) = u\}$, can be obtained for $u > 0$ as*

$$P\{\hat{B}_i^n(t_0) = u\} = P\{\Phi_0 \geq n\} \sum_{x=1}^{S_0-n+1} P\{\Psi^{n-1} = x\} \frac{\lambda_i}{\lambda_0} P\{Y_i = u + x\}, \quad (15)$$

and for $u = 0$ as

$$P\{\hat{B}_i^n(t_0) = 0\} = P\{\Phi_0 \geq n\} \sum_{x=1}^{S_0-n+1} P\{\Psi^{n-1} = x\} \left(\frac{\lambda_i}{\lambda_0} P\{Y_i = x\} + \frac{\lambda_{ic}}{\lambda_0} P\{Y_{ic} \geq x\} \right). \quad (16)$$

Turning to the retailer orders that are completely backordered, we know that after the nominal inventory position has reached zero for the n^{th} 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 n^{th} 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 n^{th} retailer order has arrived to the warehouse, given $\Phi_0 \geq n$.

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:

$$P\{Z_i^m = a\} = \binom{m}{a} \left(\frac{\lambda_i}{\lambda_0}\right)^a \left(\frac{\lambda_0 - \lambda_i}{\lambda_0}\right)^{m-a}. \quad (17)$$

Lemma 3. *The probability of v complete backorders for retailer i given that the n^{th} 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*

$$P\{\check{B}_{i,n}(t_0) = v\} = \sum_{a=1}^v P\{\check{\Phi}_{i,n} = a\} P\{Y_i^a = v\}, \quad (18)$$

where

$$P\{\check{\Phi}_{i,n} = a\} = \sum_{m=a}^{\infty} \frac{P\{\Phi_0 = n + m\}}{P\{\Phi_0 \geq n\}} P\{Z_i^m = a\} \quad (19)$$

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

$$p_0 = \sum_{n=0}^{S_0-1} P\{\Phi_0 = n\} P\{\Psi^n > 0\}. \quad (20)$$

The pmf of the amount of backordered units for $S_0 > 0$ can now be obtained from Proposition 4.

Proposition 4. *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*

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

and for $r = 0$ as

$$P\{B_i(t_0) = 0 | IP_0(t_0 - L_0) = S_0\} = p_0 + \sum_{n=1}^{S_0} P\{\hat{B}_i^n(t_0) = 0\} P\{\check{B}_{i,n}(t_0) = 0\}. \quad (22)$$

Remark 1. In case of Poisson demand there are no partial backorders at the central warehouse and the analysis is equivalent to the binomial disaggregation approach in Simon (1971) and Graves (1985).

Remark 2. The expected amount of backorders is not always proportional to the demand per time unit resulting in different average waiting times per time 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 (consider for instance two retailers, one facing unit demand and one facing larger demand sizes).

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 3. 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 $\tilde{B}_0(t_0)$ (which by definition equals $-S_0$) 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 that there are r units backordered at the central warehouse for retailer i at time t_0 , when $S_0 \leq 0$, $P\{B_i(t_0) = r\}$, can thus be obtained as a convolution of $\tilde{B}_i(t_0)$ and $\check{B}_i(t_0)$:

$$P\{B_i(t_0) = r | IP_0(t_0 - L_0) = S_0\} = \sum_{u=0}^{\min(r, -S_0)} P\{\tilde{B}_i(t_0) = u\}P\{\check{B}_i(t_0) = r - u\}, \quad S_0 \leq 0. \quad (23)$$

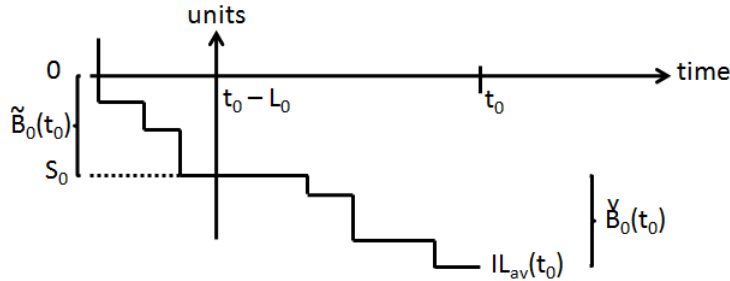


Figure 3: Illustration of a possible sample path for the backorders at the central warehouse at t_0 for $S_0 \leq 0$

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

$$P\{\check{B}_i(t_0) = v\} = P\{D_i(t_0 - L_0, t_0) = v\}. \quad (24)$$

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 (23) simplifies to $P\{B_i(t_0) = r | IP_0(t_0 - L_0) = 0\} = P\{\tilde{B}_i(t_0) = r\}$. 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 m^{th} retailer order before $t_0 - L_0$ (i.e., counting backwards from $t_0 - L_0$, the m^{th} 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

$$P\{\tilde{B}_i(t_0) = u\} = \begin{cases} \sum_{m=1}^{S_0} P\{\tilde{B}_i^{-m}(t_0) = u\}, & S_0 < 0 \\ 1, & u = 0 \text{ and } S_0 = 0, \\ 0, & \text{otherwise} \end{cases} \quad (25)$$

where $P\{\tilde{B}_i^{-m}(t_0) = u\}$ can be determined from Lemma 5.

Lemma 5. *The probability for u backordered units at t_0 , ordered by retailer i before time $t_0 - L_0$, when the m^{th} customer order before $t_0 - L_0$ is the first order to be backordered, is*

$$P\{\tilde{B}_i^{-m} = u\} = \sum_{x=S_0+m-1}^{-1} \frac{\lambda_{i^c}}{\lambda_0} P\{Y_{i^c} \geq -x\} P\{V_i^{m-1} = u \text{ and } \Psi^{-(m-1)} = x\} \\ + \frac{\lambda_i}{\lambda_0} P\{Y_i \geq -x\} P\{V_i^{m-1} = u + x \text{ and } \Psi^{-(m-1)} = x\}, \quad (26)$$

where

$$P\{V_i^{m-1} = u \text{ and } \Psi^{-(m-1)} = x\} = \sum_{a=0}^{m-1} P\{Z_i^{m-1} = a\} P\{Y_i^a = u\} P\{Y_{i^c}^{m-1-a} = x - S_0 - u\}. \quad (27)$$

$P\{B_i(t_0) = r | IP_0(t_0 - L_0) = S_0\}$ for $S_0 \leq 0$ can now be determined from (23), (24), (25) and Lemma 5.

3.4 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 retailer 3 the second. Each retailer face compound Poisson demand with a logarithmic compounding distribution, i.e., $P\{Y_i = y\} = -\alpha_i^y / (\ln(1 - \alpha_i)y)$ 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 1. As seen in Section 3.2, to analyze retailers i ($i = 1, 2, 3$) we need to determine

Table 1: Parameter values for illustrative example

$E[D_i] = \{1, 1, 1\}$	$R_0 = -2$	$T_k = \{0.5, 1\}$	$L_i = \{0.5, 0.5, 1, 0.5\}$	$\beta_i = \{10, 10, 10\}$
$\rho_i = \{4, 2, 1.5\}$	$Q_0 = 5$	$S_i = \{4, 4, 4\}$	$h_i = \{1, 1, 1, 1\}$	$\omega_k = \{2, 2\}$

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 2.

Table 2: 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 (8), (10) and (11) respectively. Calculating the expected stock on hand at the central warehouse using (3), (4) and (5), 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 3.

Table 3: 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}

¹The Standard deviations of the simulated results were < 0.001 .

4 Optimization

In this Section we present a method for optimizing the system parameters R_0 , \mathbf{S} and \mathbf{T} both in systems with backorder costs and in systems with fill rate constraints. Our approach is based on the analysis in Section 3 and we presume that the order quantity Q_0 is given by the outside supplier/manufacturer. The optimization is explained for the single-item case and is extended to the multi-item case in Section 5. The objective is to minimize the total cost function (2) or the total expected shipment and holding costs if fill rate constraints are used. We know from Section 3.1 that the stock on hand, and therefore also the holding costs, at the central warehouse can be separated into; the *availability stock*, which depends on R_0 but is independent of \mathbf{T} and \mathbf{S} , and the *consolidation stock* that depends on \mathbf{T} but is independent of R_0 and \mathbf{S} . Let:

$AC(R_0)$: The holding cost for the *availability stock* at the central warehouse.

$CC_k(T_k)$: The holding cost for the warehouse *consolidation stock* allocated to retailer group k .

$SC_k(T_k)$: The shipment costs for shipments scheduled for retailer group k .

$RC_i(R_0, S_i, T_k)$: The holding and backorder costs (when applicable) at retailer i in retailer group k .

TC_k : The total costs directly related to retailer group k , i.e. $CC_k(T_k) + SC_k(T_k) + \sum_{i \in N_k} RC_i(R_0, S_i, T_k)$.

\overline{TC} : The lowest total expected cost of the system that is known.

Expressing the expected total cost in the components above renders

$$\begin{aligned}
TC(R_0, \mathbf{S}, \mathbf{T}) &= AC(R_0) + \sum_{k \in K} \left[CC_k(T_k) + SC_k(T_k) + \sum_{i \in N_k} RC_i(R_0, S_i, T_k) \right] \\
&= AC(R_0) + \sum_{k \in K} TC_k(R_0, \mathbf{S}, T_k).
\end{aligned} \tag{28}$$

We can show by example that the total cost function is not jointly convex in R_0 , \mathbf{S} and \mathbf{T} . The proposed optimization method is therefore based on bounding R_0 and $T_k \forall k$ using Proposition 7

and Proposition 8 respectively, and searching this bounded region using the convexity property of the retailer order-up-to levels specified in Proposition 6.

Proposition 6. *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 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 determine a good initial solution (a close to optimal initial solution will provide tighter bounds). We find such a solution by fixing the shipment intervals to some reasonable values $\overline{\mathbf{T}}$, i.e. $T_k = \overline{T}_k \forall k$, and then optimizing R_0 and \mathbf{S} . Good initial values for the shipment intervals $\overline{T}_k \forall k$ may be determined from the Economic Order Interval (EOI) heuristic, based on a deterministic demand assumption, presented in Marklund (2011):

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

For these shipment intervals, the optimization of R_0 and \mathbf{S} is carried out by searching the possible values of R_0 starting from $R_0^l = -Q$ (this bound is known from previous research, see for example Marklund 2011 and Axsäter 1998), and using the convexity property of the retailer costs from Proposition 6 to optimize $S_i \forall i$ for each value of R_0 . We stop increasing R_0 when the upper bound R_0^u , obtained from Proposition 7, is reached. Note that \overline{TC} (by definition) is updated concurrently during the search as better solutions (i.e. with lower expected costs) are found. This means that R_0^u is also updated and becomes tighter as the quality of \overline{TC} improves.

Proposition 7. *An upper bound for the optimal reorder point at the central warehouse, R_0^u , is obtained for the smallest value of R_0 for which*

$$AC(R_0) \geq \overline{TC} - \sum_{k \in K} TC_k^l. \quad (30)$$

For systems with fill rate constraints at the retailers we search $R_0 \in [R_0^l, R_0^u]$, and for each R_0 we determine the smallest S_i satisfying the fill rate constraints, as we know that the costs are increasing

for higher values of S_i .

Turning to the optimization of the shipment intervals lower and upper bounds for the optimal T_k $\forall k \in K$ are provided by Proposition 8.

Proposition 8. *For the optimal shipment interval of retailer group k , a lower bound is obtained by*

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

and an upper bound by

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

The optimal solution with respect to \mathbf{T} , R_0 and \mathbf{S} can now be obtained by searching all combinations of T_k within the bounded region, 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. 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 $R_0 \in [R_0^l, R_0^u]$.

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 9 provides such a bound which is applicable both for systems with backorder costs and fill rate constraints.

Lemma 9. *A lower bound for the costs directly related to retailer group k , $TC_k = CC_k(T_k) + SC_k(T_k) + \sum_{i \in N_k} RC_i(R_0, S_i, T_k)$ for all R_0 , can be obtained as*

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

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 10 provides tighter bounds for TC_k^l in backorder cost systems.

Lemma 10. *A lower bound for the costs directly related to retailer group k , $TC_k = CC_k(T_k) + SC_k(T_k) + \sum_{i \in N_k} RC_i(R_0, S_i, T_k)$ 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 (infinite stock):*

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

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

5 Multi-Item Systems

In this section we extend 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 group of items jointly distributed to a group of retailers. In principle there can be different groups associated with different items, but evidently $K \leq N \times J$. 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 ($R_{0,1}, \dots, R_{0,J}$)

$S_{i,j}$: 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 ($S_{1,j}, \dots, S_{N,j}$)

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 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 possible backorder costs for item j are independent of the other items, and can therefore be determined as in the single-item case. The total cost function for the multi-item system with backorder costs is

$$TC_M = \sum_{j \in J} \left[HC_0(R_{0,j}, \mathbf{S}_j, \mathbf{T}) + \sum_{i \in N} [HC_i(R_{0,j}, \mathbf{S}_j, \mathbf{T}) + BC_i(R_{0,j}, \mathbf{S}_j, \mathbf{T})] \right] + SC(\mathbf{T}). \quad (35)$$

For a multi-item system with fill rate constraints, each item j at retailer i may have an individual fill rate constraint. The fill rate of item j at retailer i , $\gamma_{i,j}$, can be determined analogously to the single-item case using (11). As before, the backorder costs are excluded from the total cost function for systems with fill rate constraints.

The optimization procedure for multi-item systems is similar to the single-item optimization with the additional restriction that all items associated with a specific consolidation group use the same shipment interval. Note that the shipment intervals are the only way in which the different items affect each other. The bounds for the warehouse reorder points need to be determined separately for each item, while T_k^l and T_k^u only need to be determined once for each consolidation group, taking all associated items into consideration. In order to determine these bounds we can still use Propositions 7 and 8, provided \overline{TC} is redefined as the lowest known total expected cost for all items, 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 determined from Lemma 9 by including the holding costs of all items in the consolidation stock.

The optimization procedure is analogous to the single-item case. First, TC_k^l is determined. Second, a close to optimal 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. With \overline{TC} in place, T_k^l and $T_k^u \forall k \in K$ can be determined. Finally, the

optimal solution can be found by searching through 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 for the system.

6 Numerical Study

In this section, we illustrate the behavior of the optimal solutions for a range of different systems in order to gain some insights. We also investigate the performance of the EOI heuristic for determining the shipment intervals (29), from Marklund (2011). According to Marklund (2011) it performs very well for Poisson demand, but it is unclear if it performs well also for systems with compound Poisson demand (with higher variance-to-mean ratio). The considered test series is obtained by varying each of the parameters $N, L_i, \rho_i, \beta_i, \omega_k, L_0$ and Q_0 between a low and a high value and investigating all combinations, creating 128 problem settings. 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 all 128 problem settings 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 4 summarizes the results in terms of average effects of varying ρ_i, β_i and ω_k 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^* . The results associated with $\rho_i = 1$ are averages across all 64 problems where $\rho_i = 1$ and analogously for all other parameters. We focus on the impact of ρ_i, β_i and ω_k as the results for the other four parameters were found to be rather intuitive. We also report the relative difference between the heuristic shipment intervals and the optimal, ΔT , and the relative increase in the associated costs, ΔC ($\Delta T = (\bar{T}_k - T_k^*)/T_k^*$,

$\Delta C = (\overline{TC} - TC^*)/TC^*$, where \overline{TC} denotes the expected cost for the system when R_0 and \mathbf{S} are optimized, and the shipment intervals are determined using the heuristic).

Table 4: Average results for the test series, for low and high values of ρ_i , β_i and ω_k

	$E[TC^*]$	$E[R_0^*]$	$E[S_i^*]$	$E[T_k^*]$	$E[\Delta C]$	$E[\Delta T]$
$\rho_i = 1$	82.02	19.11	15.08	3.24	0.12%	-0.60%
$\rho_i = 5$	141.00	20.67	23.75	3.05	0.16%	6.57%
$\beta_i = 10$	93.66	18.19	15.88	3.30	0.08%	-2.27%
$\beta_i = 100$	129.37	21.59	22.94	2.99	0.21%	8.24%
$\omega_k = 10$	81.96	21.25	15.97	1.54	0.13%	1.90%
$\omega_k = 100$	141.07	18.53	22.85	4.75	0.15%	4.07%
All problems	111.51	19.89	19.41	3.15	0.14%	2.99%

Focusing on the behavior of the optimal solutions, Table 4 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. But we also see an increase in the average reorder point at the central warehouse from 18.19 to 21.59 and a quite significant decrease in the average shipment intervals (from 3.30 to 2.99), increasing the flexibility.

Turning to the shipment costs, we can see from Table 4 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 (longer lead times to the retailers mean that it takes longer to prevent a shortage at the retailer by an order from 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 (In our examples, we could see that the total stock at the central warehouse increased).

Finally, the results in Table 4 illustrate that using the EOI heuristic for determining the shipment intervals seem to work quite well 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 quite 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 we see that $E[\Delta C] = 0.21\%$. 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).

Looking at the computational aspects, the optimization times were between 0.2 and 120 minutes per problem setting on a Dell Latitude 6400 personal computer. The parameters that affect the computational time the most are $\lambda_0 L_0$, N and K . An observation for shortening the computation times is that the fairly time consuming calculations of the backorder distributions only needs to be computed once for each value of R_0 . The suggested upper bound on R_0 , R_0^u , is also found to be rather tight. This is nice as the calculations of the backorder distributions are increasing for high values of $R_0 + Q_0$. In our study, on average R_0^u was only 7.4 units higher than the optimal value for R_0 . The bounds on the shipment interval, T_k^l and T_k^u are looser, $(T_k^u - T_k^l)$ was on average 25.74 time units (compared to T^* which was on average 3.15 as seen in Table 4). However, the bounds on the shipment intervals have a smaller impact on the optimization times than R_0^u .

7 Summary and Concluding Remarks

In this paper, we provide a method for exact analysis of inventory levels, fill rates and expected costs in one warehouse multi-retailer inventory systems with time based shipment consolidation and compound Poisson demand. Furthermore, we provide an optimization procedure based on bounds

for the optimal decision variables; the warehouse reorder point, the retailer base-stock levels, and the shipment intervals. The suggested method is applicable for both single- and multi-item systems with either shortage costs or fill rate constraints at the retailers.

The numerical results show how the different control parameters are influenced as the system parameters change. The shipment intervals are primarily used to control against changes in the shipment costs, but we can also see a tendency to shorten the shipment intervals for systems with high variability in the demand and for systems with high backorder costs.

The exact approach presented in this paper can be computationally challenging to apply to large systems. Thus, one direction for future research is to consider new, computationally more efficient, approaches for cost analysis and optimization. In addition, we believe that our analysis provides a good foundation for future research on both exact analysis and more accurate heuristic methods for solving similar divergent networks with other characteristics, e.g. N -echelon distribution models, periodic replenishments to the central warehouse or demand with other distributions than compound Poisson.

8 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. Boston: Kluwer Academic Publishers.
- 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.
- 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., F. Mutlu, C.Y. Lee. 2006. A comparison of outbound dispatch policies for integrated inventory and transportation decisions. *European Journal of Operational Research* **171** 1094–1112.
- 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.
- Dogru, M.K. 2006. Optimal Control of One-Warehouse Multi-Retailer Systems: An Assessment of the Balance Assumption. Doctoral Thesis, Beta Research School for Operations Management and Logistics, Technische Universiteit Eindhoven.
- Federgruen, A., P. Zipkin. 1984a. Approximations of Dynamic Multilocation Production and Inventory Problems. *Management Science* **30** 69–84.
- Federgruen, A., P. Zipkin. 1984b. Allocation Policies and Cost Approximations for Multilocation Inventory Systems. *Naval Research Logistics Quarterly* **31** 97–129.
- 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.
- 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.
- Gürbüz, M., K. Moınzadeh, 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., J. Marklund. 2011. Evaluation of stock allocation policies in a divergent system with shipment consolidation. *European Journal of Operational Research* **211** 289–309.
- 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.
- 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.
- 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.
- Moınzadeh K. 2002. A Multi-Echelon Inventory System with Information Exchange, *Management Science* **48** 414–426.
- Mutlu, F., S. Çetinkaya. 2010. An integrated model for stock replenishment and shipment scheduling under common carrier dispatch costs. *Transportation Research Part E* **46** 844–854.
- 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.H., S.X. Zhou. 2012. Optimizing Replenishment Intervals for Two-echelon Distribution Systems with Stochastic Demand. Working paper, Fuqua School of Business, Duke University, North Carolina, USA.
- 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.
- Verrijdt J.H.C.M., A.G. de Kok, 1996. Distribution Planning for Divergent Depotless Two-Echelon Network under Service Constraints. *European Journal of Operational Research* **89** 341–354.
- 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.

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 10,

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

we go through the following steps:

1. Determine a near optimal shipment interval for retailer group k , \bar{T}_k , using the heuristic provided by Marklund(2011) in (29)
2. Given \bar{T}_k , determine near optimal reorder points for each retailer \bar{S}_i by optimizing $S_i \forall i \in N_k$ using the convexity property in Proposition 6. Note that $B_i(t_0) = 0$ implies no delays due to stockouts at the central warehouse. This means that $S_i \forall i \in N_k$ are optimized to minimize $TC_k = CC_k(\bar{T}_k) + SC_k(\bar{T}_k) + \sum_{i \in N_k} \lim_{r \rightarrow \infty} RC_i(R_0 = r, S_i, \bar{T}_k)$.
3. Calculate an initial total cost for the costs directly related to retailer group k , $\overline{TC}_k = TC_k(\bar{T}_k, \bar{S}_i)$.
4. Obtain upper and lower bounds for T_k for this optimization using the convexity of $(CC_k + SC_k)(T_k)$ in T_k proven in Proposition 8. We obtain these bounds for the two values for T_k that satisfies

$$(CC_k + SC_k)(T_k) = \overline{TC}_k. \quad (\text{A2})$$

Using a proof analogous to that in Proposition 8, lower and upper bounds are obtained as

$$T_k^l = \frac{\overline{TC}_k - \sqrt{\overline{TC}_k^2 - 2\omega_k \sum_{i \in N_k} h_0 \lambda_i \mu_i}}{\sum_{i \in N_k} h_0 \lambda_i \mu_i} \quad (\text{A3})$$

$$T_k^u = \frac{\overline{TC}_k + \sqrt{\overline{TC}_k^2 - 2\omega_k \sum_{i \in N_k} h_0 \lambda_i \mu_i}}{\sum_{i \in N_k} h_0 \lambda_i \mu_i} .. \quad (\text{A4})$$

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

Appendix B: Proofs

Proof of Proposition 1

Proof. The arriving process to the *consolidation stock* is equal to the departure process from the *availability stock* and it is independent of the length of the shipment interval, and when shipment leave the central warehouse. Because all unsatisfied demand is backordered, and every unit passes the consolidation stock, the expected number of units reserved by retailer i that enters the *consolidation stock* per time unit will in the long run be $\lambda_i \mu_i$.

As a consequence, we know that if the previous shipment to retailer $i \in N_k$ left at time t , the expected number of units in the *consolidation stock*, reserved by retailer i , at $t + \tau$, for any $\tau \in (0, T_k]$, is $\lambda_i \mu_i \tau$. This means that the expected amount of units reserved by retailer i in the consolidation stock will

increase linearly between two consecutive shipments from 0 to $\lambda_i \mu_i T_k$. Taking the average over time and summing over all retailers in all retailer groups renders (5). \square

Proof of Lemma 2:

Proof. Conditioned on that there are 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 n^{th} retailer order if two conditions are fulfilled: (i) The n^{th} 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 $u > 0$,

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

(15) follows from unconditioning by taking the expectation for all possible positive values for $\Psi^{n-1} = x$, $x \in [1, S_0 - n + 1]$ and all $\Phi_0 \geq n$.

In order to have 0 partial backorders for retailer i , when the n^{th} 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 n^{th} customer arrives from retailer i and demands exactly x units to move the nominal inventory position to 0, or the n^{th} customer arrives from another retailer and demands more than or equal to x units. \square

Proof of Lemma 3:

Proof. As stated before, 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 n^{th} retailer order, i.e., $\check{\Phi}_{i,n} = a$, it is clear that $P\{\check{B}_{i,n}(t_0) = v | \check{\Phi}_{i,n} = a\} = P\{Y_i^a = v\}$. (18) follows by taking the expectation over all possible outcomes of $\check{\Phi}_{i,n}$.

To arrive at (19) 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 $P\{\Phi_0 = n + m\} / P\{\Phi_0 \geq n\}$. (19) follows as an expectation over all possible values of m . \square

Proof of Proposition 4:

Proof. By definition $P\{\hat{B}_i^n(t_0) = u\}$ is the probability that the n^{th} 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 is $r - u$ complete backorders for retailer i conditioned on that the n^{th} 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 (21).

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 (22). \square

Proof of Lemma 5:

Proof. In order for $\tilde{B}_i^{-m}(t_0) = u$ there are two possible scenarios; (a) The m^{th} 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 m^{th} order originates with retailer i and is for at least $-\Psi^{-m} = -x$ units. Starting with (a) the probability that the m^{th} 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 $P\{V_i^{m-1} = u$ and $\Psi^{-(m-1)} = x\}$. Summing up for all possible values of x ($x \in [S_0 + m - 1, -1]$) generates the first part of (26).

The probability for scenario (b) is $(\lambda_i/\lambda_0)P\{Y_i \geq -x\}$. In this scenario, $-x$ units of the m^{th} 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 $P\{V_i^{m-1} = u + x$ and $\Psi^{-(m-1)} = x\}$. A summation over all possible values of x renders the second part of (26).

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 (17) we get (27). \square

Proof of Proposition 6:

Proof. The shipment cost $SC_k(T_k)$ does not depend on any S_i , and from the Section 3.1, we know that $H_0(R_0, \mathbf{S}, T_k)$ can be separated in $AC(R_0)$ and $CC_k(T_k)$, which are neither dependent on S_i . Furthermore, the holding cost and backorder cost at retailer i does not depend on order-up-to levels of other retailers. Thus, the cost function is separable in retailer order-up-to levels. Then, it is sufficient to show convexity for $RC_i(R_0, S_i, T_k)$ with respect to S_i for each retailer i . First, we define the difference function $\Delta G(s)$ as follows:

$$\Delta G(s) = RC_i(R_0, s + 1, T_k) - RC_i(R_0, s, T_k). \quad (\text{B2})$$

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

$$\begin{aligned} RC_i(R_0, s, T_k) &= (h_i + \beta_i)E[IL_i^+(s)] - \beta_i E[IL_i(s)], \\ &= (h_i + \beta_i) \frac{1}{T_k} \sum_{j=1}^s \sum_{r=0}^{s-j} j P\{B_i(t_0) = r\} \int_0^{T_k} P\{D_i(L_i + x) = s - j - r\} dx, \\ &\quad - \beta_i (s - E[B_i(t_0)] - \lambda_i \mu_i L_i). \end{aligned} \quad (\text{B3})$$

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. Then, we derive $RC_i(R_0, s + 1, T_k)$ in terms of $RC_i(R_0, s, T_k)$:

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

$$\begin{aligned}
= & (h_i + \beta_i) \frac{1}{T_k} \sum_{z=0}^s \sum_{r=0}^{s-z} z P\{B_i(t_0) = r\} \int_0^{T_k} P\{D_i(L_i + x) = s - z - r\} dx \\
& + (h_i + \beta_i) \frac{1}{T_k} \sum_{z=0}^s \sum_{r=0}^{s-z} P\{B_i(t_0) = r\} \int_0^{T_k} P\{D_i(L_i + x) = s - z - r\} dx \\
& - \beta_i(s - E[B_i(t_0)] - \lambda_i \mu_i L_i) - \beta_i, \\
= & (h_i + \beta_i) \frac{1}{T_k} \sum_{z=0}^s \sum_{r=0}^{s-z} P\{B_i(t_0) = r\} \int_0^{T_k} P\{D_i(L_i + x) = s - z - r\} dx \\
& - \beta_i + RC_i(R_0, s, T_k).
\end{aligned} \tag{B5}$$

As a result the first order difference is equal to:

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

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

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

□

Proof of Proposition 7:

Proof. It is clear from (4) that $AC(R_0) = h_0 E[IL_{av}^+]$ is increasing in R_0 for all values of $R_0 \geq -Q_0$ and that the total cost $TC(R_0, \mathbf{S}, \mathbf{T}) = AC(R_0) + \sum_{k \in K} TC_k(R_0, \mathbf{S}, T_k) \geq AC(R_0) + \sum_{k \in K} TC_k^l$. With $R_0^u = \min\{R_0 : AC(R_0) + \sum_{k \in K} TC_k^l \geq \overline{TC}\}$ 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. □

Proof of Proposition 8:

Proof. Note that the costs directly related to the shipment interval T_k ,

$$(CC_k + SC_k)(T_k) = \frac{1}{2} \sum_{i \in N_k} h_0 \lambda_i \mu_i T_k + \frac{\omega_k}{T_k}, \quad (\text{B8})$$

are strictly convex in T_k as (for $T_k > 0$)

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

Note also that

$$\begin{aligned} TC(R_0, \mathbf{S}, \mathbf{T}) &= AC(R_0) + \sum_{\kappa \in K} TC_\kappa(R_0, \mathbf{S}, T_\kappa) \\ &\geq TC_k(R_0, \mathbf{S}, T_k) + \sum_{\kappa \in K \setminus \{k\}} TC_\kappa(R_0, \mathbf{S}, T_\kappa) \geq (CC_k + SC_k)(T_k) + \sum_{\kappa \in K \setminus \{k\}} TC_\kappa^l. \end{aligned} \quad (\text{B10})$$

It follows that no T_k satisfying

$$(CC_k + SC_k)(T_k) \geq \overline{TC} - \sum_{\kappa \in K \setminus \{k\}} TC_\kappa^l. \quad (\text{B11})$$

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

$$\frac{1}{2} \sum_{i \in N_k} h_0 \lambda_i \mu_i T_k + \frac{\omega_k}{T_k} = \overline{TC} - \sum_{\kappa \neq k} TC_\kappa^l \quad (\text{B12})$$

with respect to T_k . □

Proof of Lemma 9:

Proof. From Proposition 8 we know that $(CC_k + SC_k)(T_k)$ is convex in T_k . Hence, the shipment interval $T_{k,l}$ that minimizes $(CC_k + SC_k)(T_k)$ is obtained from the first order optimality condition

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

which renders

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

(33) follows from

$$TC_k(T_k) \geq (CC_k + SC_k)(T_k) \geq (CC_k + SC_k)(T_{k,l}) = \sqrt{2h_0\omega_k \sum_{i \in N_k} \lambda_i \mu_i}. \quad (\text{B15})$$

□

Proof of Lemma 10:

Proof. Note first that neither $CC_k(T_k)$ nor $SC_k(T_k)$ depend on R_0 . Moreover, the retailer costs, $RC_i(R_0, S_i, T_k)$, depend on R_0 solely through the backorder distribution, $B_i(t_0)$. From (7) 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

$$\begin{aligned} P\{IL_i(t_0 + L_i + t) = j\} &= P\{b_i + D_i(t_0, t_0 + L_i + t) = S_i - j\} \\ &= P\{(b_i + 1) + D_i(t_0, t_0 + L_i + t) = (S_i + 1) - j\}, \end{aligned} \quad (\text{B16})$$

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 , $\{1, 2, \dots, N_k\}$. Now, let \mathbf{S}_{N_k} denote the vector of all order-up-to levels within retailer group k , $\{S_1, \dots, S_{N_k}\}$, and $\mathbf{B}_{N_k}(t_0)$ denote the vector of backordered units to each retailer in retailer group k at t_0 , $\{B_1(t_0), \dots, B_{N_k}(t_0)\}$. 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:

$$TC_k(R_0, \mathbf{S}_{N_k}, T_k) = \sum_{b_1=0}^{\infty} \dots \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, \dots, b_k\}\} TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{b_1, \dots, b_k\}). \quad (\text{B18})$$

This gives us for any value of R_0

$$\begin{aligned} TC_k(R_0, \mathbf{S}, T_k) &\geq \min_{T_k, \mathbf{S}_{N_k}} [TC_k(R_0, \mathbf{S}, T_k)] \\ &= \min_{T_k, \mathbf{S}_{N_k}} \left[\sum_{b_1=0}^{\infty} \dots \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, \dots, b_k\}\} TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{b_1, \dots, b_k\}) \right] \\ &\geq \sum_{b_1=0}^{\infty} \dots \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, \dots, b_k\}\} \min_{T_k, \mathbf{S}_{N_k}} [TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{b_1, \dots, b_k\})] \\ &= \sum_{b_1=0}^{\infty} \dots \sum_{b_k=0}^{\infty} P\{\mathbf{B}_{N_k}(t_0) = \{b_1, \dots, b_k\}\} \min_{T_k, \mathbf{S}_{N_k}} [TC_k(\mathbf{S}_{N_k}, T_k | \mathbf{B}_{N_k}(t_0) = \{0, \dots, 0\})] \\ &= \min_{T_k, S_i \forall i \in N_k} (TC_k | B_i(t_0) = 0, \forall i \in N_k), \end{aligned} \quad (\text{B19})$$

which proves (34). Note that the next to last equality follows from (B17) and the fact that the retailer costs are the only costs affected by the backorder distribution. The last equality follows directly as probabilities must sum to 1. □

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 policies. 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.

Keywords: Inventory, Multi-echelon, Batch ordering, Stochastic demand, Delivery 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
\mathbf{N}	=	$\{1,2,\dots,N\}$, 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, \dots, 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 $Q_0 \geq \max(Q_1, \dots, Q_N)$. 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 $Q_0 \geq \max(Q_1, \dots, Q_N) - q$, 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 $\mathbf{R} = (R_0, \dots, R_N)$ that minimize expected total system costs for each of the three different delivery policies. The expected total costs are given by

$$TC(\mathbf{R}) = h_0(E[IL_0^+] + \sum_{i=1}^N E[O_i]) + \sum_{i=1}^N \theta_i S_i + \sum_{i=1}^N (h_i E[IL_i^+] + b_i E[IL_i^-]). \quad (1)$$

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 $q - nQ_0 \leq IL_0(\tau_z) \leq Q_i - q - nQ_0$ ($n = 0, 1, 2, \dots$) 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 n^{th} unit in the u units considered will satisfy the $(R_i + n)^{\text{th}}$ customer demand after τ_z at retailer i . If $R_i + n \leq 0$, the n^{th} unit will satisfy the $(R_i + n + 1)^{\text{th}}$ 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 n^{th} 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 n^{th} unit is demanded is Erlang distributed. Let

$$\begin{aligned} \Gamma_i(k) &= \text{stochastic time for } k \text{ customer demand arrivals at retailer } i, \Gamma_i(k) \in \text{Erlang}(k, \lambda_i) \\ &\text{with probability density function } f_{\Gamma_i(k)}(x) = \lambda^k x^{k-1} e^{-\lambda x} / (k-1)! \quad x \geq 0. \\ g_i(R_i, \delta, n) &= \text{expected holding - and backorder cost at retailer } i \text{ for the } n^{\text{th}} \text{ unit placed on} \\ &\text{hold, arriving at retailer } i \text{ in } \delta \text{ time units} \end{aligned}$$

It follows that

$$\Delta C_i(R_i, t_1, t_2, u) = h_0(t_2 - t_1)u + \sum_{n=1}^u g_i(R_i, L_i + t_2, n) - g_i(R_i, L_i + t_1, n). \quad (2)$$

The n^{th} 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

$$\begin{aligned} g_i(R_i, \delta, n) &= h_i E\left[(\Gamma_i(R_i + n) - \delta)^+\right] + b_i E\left[(\Gamma_i(R_i + n) - \delta)^-\right] \\ &= (h_i + b_i)E\left[(\Gamma_i(R_i + n) - \delta)^+\right] + b_i E\left[(\delta - \Gamma_i(R_i + n))\right] \\ &= \frac{(h_i + b_i)e^{-\lambda_i \delta}}{\lambda_i} \sum_{j=0}^{R_i+n} (R_i + n - j) \frac{(\lambda_i \delta)^j}{j!} + \frac{b_i}{\lambda_i} (\lambda_i \delta - R_i - n), \end{aligned} \quad (3)$$

and for $R_i + n \leq 0$

$$g_i(R_i, \delta, n) = b_i \delta. \quad (4)$$

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

$$\Delta C_i(R_i, t_1, t_2, u) \leq \theta_i. \quad (5)$$

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, \mathbf{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 \mathbf{R} , we need to determine $E[IL_0^+]$, $E[O_i]$, S_i , $E[IL_i^+]$ and $E[IL_i^-]$ ($i = 1, \dots, N$). 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

$$\begin{aligned} E[IL_i^+] &= \sum_{m=1}^{R_i+Q_i} mP(IL_i = m) \\ E[IL_i^-] &= \sum_{m=1}^{\infty} mP(IL_i = -m). \end{aligned} \tag{6}$$

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 $[R_i+1, R_i+2, \dots, R_i+Q_i]$, and the central warehouse inventory position is uniform on $[R_0+q, R_0+2q, \dots, R_0+Q_0]$ (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

$$\begin{aligned}
D_i(\tau_1, \tau_2) &= \text{customer demand at retailer } i \text{ in the time interval } [\tau_1, \tau_2] \text{ (Poisson distributed)} \\
D_0(\tau_1, \tau_2) &= \text{demand from all retailers to the central warehouse in the time interval } [\tau_1, \tau_2] \\
D_{0,i}(\tau_1, \tau_2, a, d) &= \text{demand from retailer } i \text{ to the central warehouse in the time interval } [\tau_1, \tau_2], \\
&\quad \text{given that } IP_i(\tau_1) = a \text{ and } D_i(\tau_1, \tau_2) = d \\
\text{mod}_{R,Q}(z) &= z + nQ, \text{ where } n \text{ is the integer such that } R + 1 \leq z + nQ \leq R + Q.
\end{aligned}$$

$D_{0,i}(\tau_1, \tau_2, a, d)$ is easily obtained as nQ_i , where n is the integer such that $R + 1 \leq a - d + nQ \leq R + Q$.

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

$$IL_i(t + L_0 + L_i) = IP_i(t + L_0) - B_i(t + L_0) - O_i(t + L_0) - D_i(t + L_0, t + L_0 + L_i). \quad (7)$$

For notational convenience we assume that $t = 0$ and hereafter refer to this arbitrary time as time zero.

That is,

$$IL_i(L_0 + L_i) = IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i). \quad (8)$$

$D(L_0, L_0 + L_i)$ 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 $0 \leq \tau \leq L_0$. 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,

$$\Psi_0(\tau) = IP_0(0) - D_0(0, \tau). \quad (9)$$

$\Psi_0(\tau)$ 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 ($0 \leq \tau \leq L_0$):
- 3) If $\Psi_0(\tau) \geq 0$, no retailer orders occurring before time τ will be backordered or on hold at the central warehouse at time L_0 .
- 4) If $\Psi_0(\tau) \leq 0$, 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 $\Psi_0(\tau) = u$ ($q \leq u \leq Q_i - q$) 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 ,

$$\mathbf{V}(\tau) = [\Psi_0(\tau), IP_1(\tau), \dots, IP_N(\tau)]. \quad (10)$$

The analysis is based on calculating the probabilities for state transitions of the vector $\mathbf{V}(\tau)$. That is, given a state $\mathbf{V}(\tau_1) = \mathbf{v} = [v_0, v_1, \dots, v_N]$ at time τ_1 , we can calculate the probability of state $\mathbf{V}(\tau_2) = \mathbf{w} = [w_0, w_1, \dots, w_N]$ 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

$$P(\mathbf{V}(\tau_2) = \mathbf{w} | \mathbf{V}(\tau_1) = \mathbf{v}) = \sum_{d_1=0}^{\infty} \sum_{d_2=0}^{\infty} \dots \sum_{d_N=0}^{\infty} \chi \prod_{i=1}^N P(D_i(\tau_1, \tau_2) = d_i), \quad (11)$$

where

$$\chi = \begin{cases} 1; & v_0 - \sum_{i=1}^N D_{0,i}(\tau_1, \tau_2, v_i, d_i) = w_0, \quad \text{mod}_{R_1, Q_1}(v_1 - d_1) = w_1, \dots, \text{mod}_{R_N, Q_N}(v_N - d_N) = w_N \\ 0; & \text{otherwise.} \end{cases}$$

Proof

Because each retailer faces an independent Poisson process, the probability of a demand realization d_1, d_2, \dots, d_N in the time interval $[\tau_1, \tau_2]$ is $\prod_{i=1}^N P(D_i(\tau_1, \tau_2) = d_i)$. 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 $\text{mod}_{R_i, Q_i}(v_i - d_i)$ 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, $\mathbf{V}(0) = \mathbf{a} = [a_0, a_1, \dots, a_N]$. 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

$$P(IL_i(L_0 + L_i) = m) = \frac{q}{\prod_{j=0}^N Q_j} \sum_{\mathbf{a} \in \mathbf{A}} P(IL_i(L_0 + L_i) = m | \mathbf{V}(0) = \mathbf{a}), \quad (12)$$

where the state space \mathbf{A} contains all possible state vectors at time zero. That is,

$$\mathbf{A} = \{\mathbf{a} \in \mathbb{Z}^{N+1} \mid a_0 = R_0 + q, R_0 + 2q, \dots, R_0 + Q_0; \\ a_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N}\}. \quad (13)$$

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

$$P(IL_i(L_0 + L_i) = m | \mathbf{V}(0) = \mathbf{a}) \\ = P(IL_i(L_0 + L_i) = m, \text{I}) + P(IL_i(L_0 + L_i) = m, \text{II}) + P(IL_i(L_0 + L_i) = m, \text{III}). \quad (14)$$

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

$$\boldsymbol{\Omega} = \{\boldsymbol{\omega} \in \mathbb{Z}^{N+1} \mid \omega_0 = q, 2q, \dots, R_0 + Q_0; \\ \omega_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N}\}. \quad (15)$$

Lemma 3 provides the key to obtaining the distribution of the inventory level.

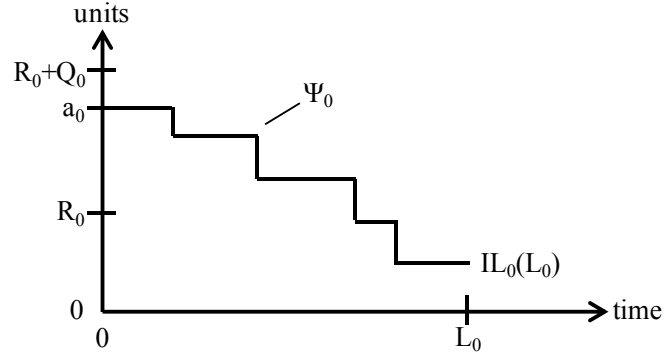


Figure 1

Lemma 3

For $\omega \in \Omega$ we have

$$P(IL_i(L_0 + L_i) = m, I | \mathbf{V}(L_0) = \omega) = P(D_i(L_0, L_0 + L_i) = \omega_i - m). \quad (16)$$

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 $\mathbf{V}(L_0) = \omega$ we have

$$\begin{aligned} IL_i(L_0 + L_i) &= IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i) \\ &= \omega_i - D_i(L_0, L_0 + L_i) \end{aligned} \quad (17)$$

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 Ω . That is, using the law of total probability we have

$$P(IL_i(L_0 + L_i) = m, I) = \sum_{\omega \in \Omega} P(D_i(L_0, L_0 + L_i) = \omega_i - m) P(\mathbf{V}(L_0) = \omega | \mathbf{V}(0) = \mathbf{a}), \quad (18)$$

where the state transition probabilities are provided in Lemma 2. Note that, because not all state transitions considered in (18) are possible, $P(\mathbf{V}(L_0) = \omega | \mathbf{V}(0) = \mathbf{a})$ 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 $\tau_w^{(-)}$ and $\tau_w^{(+)}$ to denote the time just before and just after time τ_w , respectively. That is, $\tau_w^{(-)} < \tau_w < \tau_w^{(+)}$, where $\tau_w^{(+)} = \tau_w^{(-)} + d\tau_w$, $d\tau_w \rightarrow 0$. Figure 2 depicts a possible sample path of the nominal inventory position in Event II.

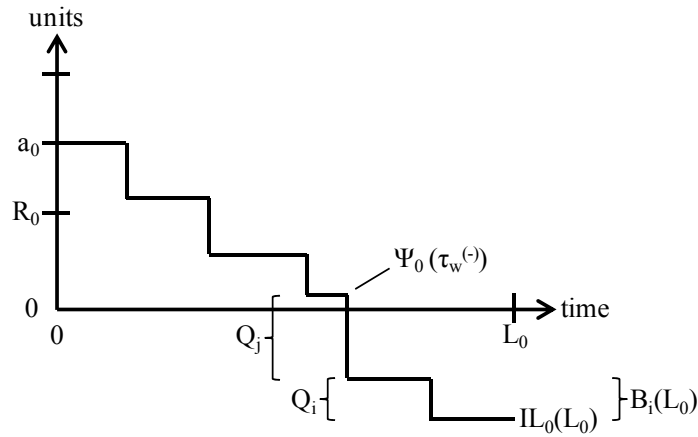


Figure 2

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

$$\begin{aligned} \mathbf{W}_j^{(-)} = \{ \mathbf{w} \in \mathbb{Z}^{N+1} \mid w_0 = q, 2q, \dots, Q_j; w_j = R_j + 1; \\ w_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus \{j\} \}. \end{aligned} \quad (19)$$

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 $j \in \mathbf{N} \setminus i$ and $\mathbf{w} \in \mathbf{W}_j^{(-)}$ we have

$$P\left(\text{IL}_i(L_0 + L_i) = m, \Pi \mid \mathbf{V}(\tau_w^{(-)}) = \mathbf{w}, D_j(\tau_w^{(-)}, \tau_w^{(+)}) = 1 \right) = P\left(D_i(\tau_w^{(+)}, L_0 + L_i) = w_i - m \right). \quad (20)$$

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 $d\tau_w \rightarrow 0$ is zero, it follows that $\text{IP}_i(\tau_w^{(-)}) = \text{IP}_i(\tau_w^{(+)}) = w_i$. Hence,

the inventory position at time L_0 must be $IP_i(L_0) = \text{mod}_{R,Q}(w_i - D_i(\tau_w^{(+)}, L_0))$
 $= w_i - D_i(\tau_w^{(+)}, L_0) + D_{0,i}(\tau_w^{(+)}, L_0, w_i, D_i(\tau_w^{(+)}, L_0))$. 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

$$\begin{aligned} IL_i(L_0 + L_i) &= IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i) \\ &= w_i - D_i(\tau_w^{(+)}, L_0) + D_{0,i}(\tau_w^{(+)}, L_0, w_i, D_i(\tau_w^{(+)}, L_0)) \\ &\quad - D_{0,i}(\tau_w^{(+)}, L_0, w_i, D_i(\tau_w^{(+)}, L_0)) - D_i(L_0, L_0 + L_i) \\ &= w_i - D_i(\tau_w^{(+)}, L_0 + L_i) \end{aligned} \quad (21)$$

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

$$\begin{aligned} P(IL_i(L_0 + L_i) = m, \Pi) &= \\ &\int_0^{L_0} \sum_{j \in \mathbf{N} \setminus i} \sum_{\mathbf{w} \in \mathbf{W}_j^{(-)}} P(D_i(\tau_w, L_0 + L_i) = w_i - m) P(\mathbf{V}(\tau_w) = \mathbf{w} \mid \mathbf{V}(0) = \mathbf{a}) \lambda_j d\tau_w. \end{aligned} \quad (22)$$

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 $k = 1, \dots, N$, 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 $j = 1, \dots, N$, 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

$$P(IL_i(L_0 + L_i) = m, III) = P(IL_i(L_0 + L_i) = m, IIIa) + P(IL_i(L_0 + L_i) = m, IIIb). \quad (23)$$

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,

$$\Psi_0(\tau) = IP_0(0) + D_0(\tau, 0). \quad (24)$$

Figure 3 depicts a possible sample path of the nominal inventory position in Event IIIa.

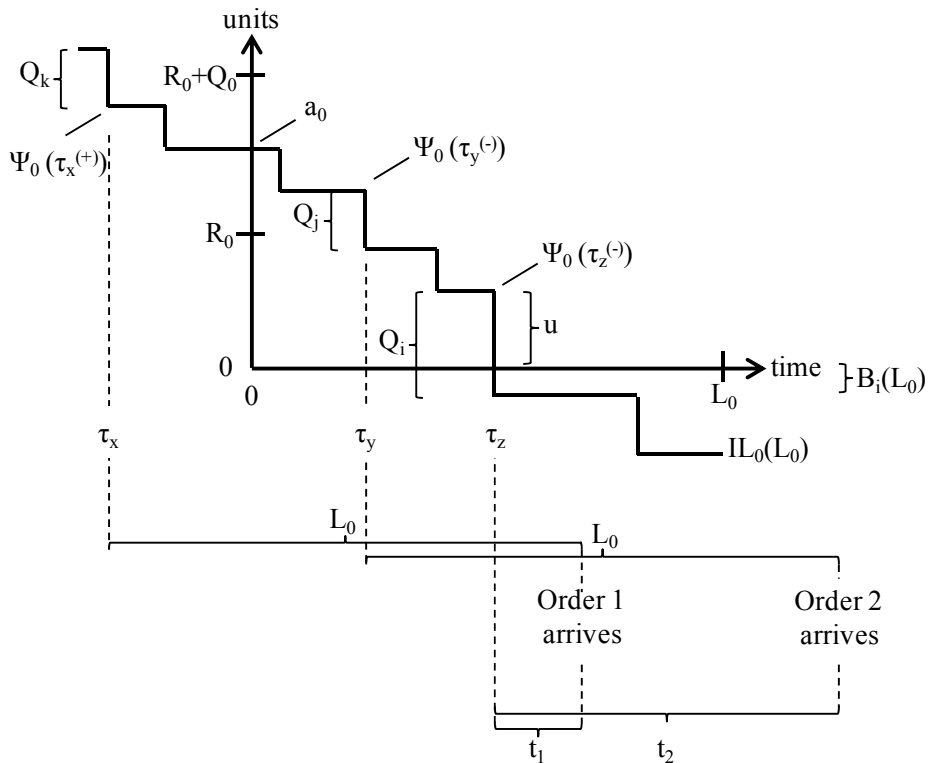


Figure 3

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 ($k = 1, 2, \dots, N$) 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

$$\begin{aligned} \mathbf{X}_k^{(+)} = \{ & \mathbf{x} \in \mathbb{Z}^{N+1} \mid x_0 = R_0 + Q_0 - Q_k + q, R_0 + Q_0 - Q_k + 2q, \dots, R_0 + Q_0; \ x_k = R_k + Q_k; \\ & x_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus \{k\} \}. \end{aligned} \quad (25)$$

At time $\tau_y^{(-)}$, a demand at retailer j ($j = 1, 2, \dots, N$) is about to trigger Order 2, but the nominal inventory position will remain positive (because we are considering Event IIIa). Hence, $\mathbf{V}(\tau_y^{(-)}) = \mathbf{y} \in \mathbf{Y}_j^{(-)}$,

$$\begin{aligned} \mathbf{Y}_j^{(-)} = \{ & \mathbf{y} \in \mathbb{Z}^{N+1} \mid y_0 = \max(R_0, Q_j) + q, \max(R_0, Q_j) + 2q, \dots, R_0 + Q_j; \ y_j = R_j + 1; \\ & y_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus \{j\} \}. \end{aligned} \quad (26)$$

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, $\mathbf{V}(\tau_y^{(+)}) = \mathbf{f}_j(\mathbf{y})$, where

$$\mathbf{f}_j(\mathbf{y}) = (y_0 - Q_j, y_1, \dots, y_j - 1 + Q_j, \dots, y_N), \quad \mathbf{y} \in \mathbf{Y}_j^{(-)}. \quad (27)$$

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 $\mathbf{V}(\tau_z^{(-)}) = \mathbf{z} \in \mathbf{Z}_i^{(-)}$,

$$\begin{aligned} \mathbf{Z}_i^{(-)} = \{ & \mathbf{z} \in \mathbb{Z}^{N+1} \mid z_0 = q, 2q, \dots, Q_i; \ z_i = R_i + 1; \\ & z_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus \{i\} \}. \end{aligned} \quad (28)$$

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

$$\begin{aligned} & \mathbb{P}(\text{II}_i(L_0 + L_i) = m, \text{IIIa} \mid D_k(\tau_x^{(-)}, \tau_x^{(+)}) = 1, \mathbf{V}(\tau_x^{(+)}) = \mathbf{x}, \\ & \quad \mathbf{V}(\tau_y^{(-)}) = \mathbf{y}, D_j(\tau_y^{(-)}, \tau_y^{(+)}) = 1, \mathbf{V}(\tau_z^{(-)}) = \mathbf{z}, D_i(\tau_z^{(-)}, \tau_z^{(+)}) = 1) = \\ & \quad \left\{ \begin{array}{l} \mathbb{P}(D_i(\tau_z^{(+)}, L_0 + L_i) = R_i - m) \quad ; \quad u < Q_i \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) \leq \theta_i \\ \mathbb{P}(D_i(\tau_z^{(+)}, L_0 + L_i) = R_i + u - m); \quad \text{otherwise} \end{array} \right. , \end{aligned} \quad (29)$$

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 $\Delta C_i(R_i, T_1, T_2, u) \leq \theta_i$. Otherwise $O_i(L_0) = 0$. In the former case (complete delivery) we therefore have

$$\begin{aligned} \Pi_i(L_0 + L_i) &= IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i) \\ &= R_i + Q_i - D_i(\tau_z^{(+)}, L_0) + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0)) \\ &\quad - \left(Q_i - u + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0)) \right) - u - D_i(L_0, L_0 + L_i) \\ &= R_i - D_i(\tau_z^{(+)}, L_0 + L_i), \end{aligned} \quad (30)$$

and in the latter case (partial delivery or if $u = Q_i$) we have

$$\begin{aligned} \Pi_i(L_0 + L_i) &= IP_i(L_0) - B_i(L_0) - O_i(L_0) - D_i(L_0, L_0 + L_i) \\ &= R_i + Q_i - D_i(\tau_z^{(+)}, L_0) + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0)) \\ &\quad - \left(Q_i - u + D_{0,i}(\tau_0, L_0, R_i + Q_i, D_i(\tau_z^{(+)}, L_0)) \right) - D_i(L_0, L_0 + L_i) \\ &= R_i + u - D_i(\tau_z^{(+)}, L_0 + L_i). \end{aligned} \quad (31)$$

■

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

$$\lambda_k d\tau_x P(\mathbf{V}(0) = \mathbf{a} | \mathbf{V}(\tau_x^{(+)}) = \mathbf{x}) P(\mathbf{V}(\tau_y^{(-)}) = \mathbf{y} | \mathbf{V}(0) = \mathbf{a}) \lambda_j d\tau_y P(\mathbf{V}(\tau_z^{(-)}) = \mathbf{z} | \mathbf{V}(0) = \mathbf{f}_j(\mathbf{y})) \lambda_i d\tau_z. \quad (32)$$

Utilizing Lemma 5 and considering all possible times, retailers and state vectors we have

$$P(\Pi L_i(L_0 + L_i) = m, \text{IIIa}) = \int_{-\infty}^0 \int_0^{L_0} \int_{\tau_y}^{L_0} \sum_{k \in \mathbb{N}} \sum_{j \in \mathbb{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\mathbf{y} \in \mathbf{Y}_j^{(-)}} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} p_i(m, \tau_x, \tau_y, \tau_z, \mathbf{z}) \pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) d\tau_z d\tau_y d\tau_x, \quad (33)$$

where $p_i()$ is given by Lemma 5,

$$p_i(m, \tau_x, \tau_y, \tau_z, \mathbf{z}) = \begin{cases} P(D_i(\tau_z, L_0 + L_i) = R_i - m) & ; u < Q_i \text{ and } \Delta C_i(R_i, t_1, t_2, u) \leq \theta_i \\ P(D_i(\tau_z, L_0 + L_i) = R_i + u - m) & ; \text{otherwise} \end{cases} \quad (34)$$

$$u = z_0, t_1 = (\tau_x + L_0 - \tau_z)^+, t_2 = \tau_y + L_0 - \tau_z,$$

and

$$\pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) = \lambda_k P(\mathbf{V}(0) = \mathbf{a} \mid \mathbf{V}(\tau_x) = \mathbf{x}) P(\mathbf{V}(\tau_y) = \mathbf{y} \mid \mathbf{V}(0) = \mathbf{a}) \lambda_j P(\mathbf{V}(\tau_z) = \mathbf{z} \mid \mathbf{V}(0) = \mathbf{f}_j(\mathbf{y})) \lambda_i \quad (35)$$

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.

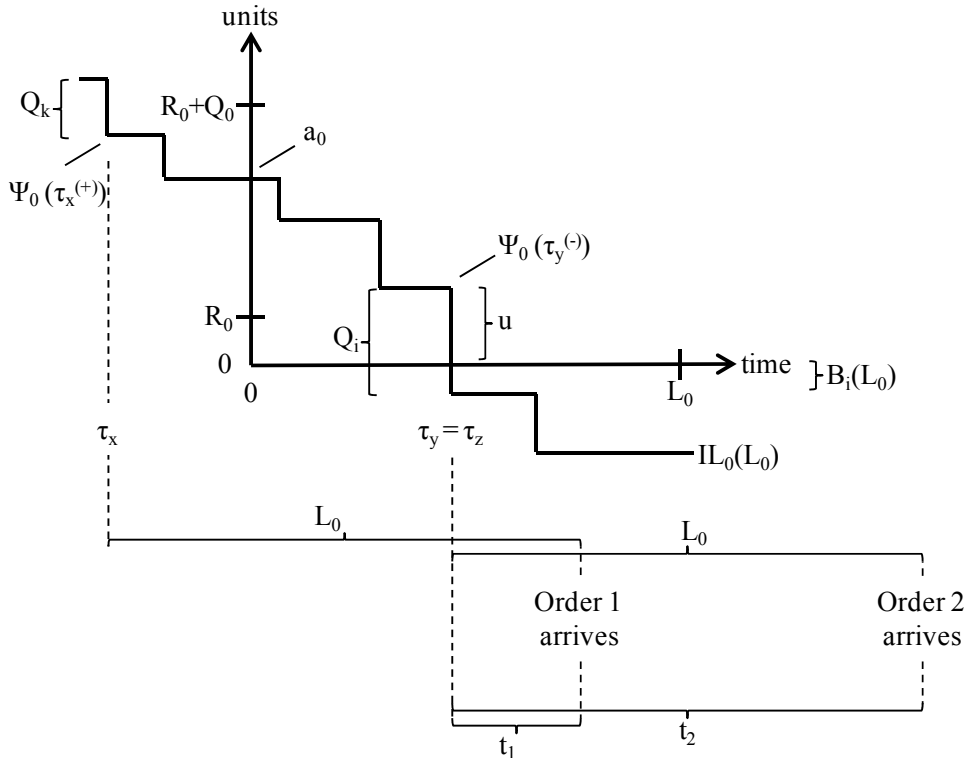


Figure 4

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 $\mathbf{Y}_j^{(-)}$ and $\mathbf{Z}_j^{(-)}$. 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

$$\begin{aligned} \tilde{\mathbf{Y}}_i^{(-)} = \{ & \tilde{\mathbf{y}} \in \mathbb{Z}^{N+1} \mid \tilde{y}_0 = R_0 + q, R_0 + 2q, \dots, Q_i; \tilde{y}_i = R_i + 1; \\ & \tilde{y}_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus i\}. \end{aligned} \quad (36)$$

Following the same analysis as above we have

$$\begin{aligned} P(\text{IIIb}) &= P(\text{II}_i(L_0 + L_i) = m, \text{IIIb}) = \\ & \int_{-\infty}^0 \int_0^{L_0} \sum_{k \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}} \tilde{p}_i(m, \tau_x, \tau_y, \tilde{\mathbf{y}}) \tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) d\tau_y d\tau_x, \end{aligned} \quad (37)$$

with

$$\begin{aligned} \tilde{p}_i(m, \tau_x, \tau_y, \tilde{\mathbf{y}}) &= \\ & \begin{cases} P(D_i(\tau_y, L_0 + L_i) = R_i - m) & ; \quad u < Q_i \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) \leq \theta_i \\ P(D_i(\tau_y, L_0 + L_i) = R_i + u - m) & ; \quad \text{otherwise} \end{cases} \\ & u = \tilde{y}_0, t_1 = (\tau_x + L_0 - \tau_y)^+, t_2 = L_0, \end{aligned} \quad (38)$$

and

$$\begin{aligned} \tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) &= \\ & \lambda_k P(\mathbf{V}(0) = \mathbf{a} \mid \mathbf{V}(\tau_x) = \mathbf{x}) P(\mathbf{V}(\tau_y) = \tilde{\mathbf{y}} \mid \mathbf{V}(0) = \mathbf{a}) \lambda_i. \end{aligned} \quad (39)$$

3.3 Determining $E[O_i]$, S_i and $E[\text{II}_0^+]$ 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, $E[\text{II}_0^+]$, are all based on the analysis in Section 3.2. Note that given the distributions of $O_i(L_0)$ and $\text{II}_0^+(L_0)$, the expected values are obtained as

$$E[O_i] = \sum_{u=q}^{Q_i-q} u P(O_i(L_0) = u), \quad (40)$$

$$E[\text{II}_0^+] = \sum_{m=q}^{R_0+Q_0} m P(\text{II}_0(L_0) = m). \quad (41)$$

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 $u = q, 2q, \dots, Q_i - q$. 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

$$P(O_i(L_0) = u) = \begin{cases} \frac{q}{N} \sum_{a \in \Lambda} P(O_i(L_0) = u, \text{IIIa}) + P(O_i(L_0) = u, \text{IIIb}); & u = q, 2q, \dots, Q_i - q \\ 0 & \text{otherwise.} \end{cases} \quad (42)$$

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 $\mathbf{V}(\tau_z^{(-)}) = \mathbf{z} \in \mathbf{Z}_i^{(-)}(\mathbf{u})$,

$$\mathbf{Z}_i^{(-)}(\mathbf{u}) = \{\mathbf{z} \in \mathbb{Z}^{N+1} \mid z_0 = u; z_i = R_i + 1; z_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus i\}, \quad (43)$$

for $u = q, 2q, \dots, Q_i - q$, and $\mathbf{Z}_i^{(-)}(\mathbf{u}) = \emptyset$ otherwise. Following the same logic as in Section 3.2.3 yields

$$P(O_i(L_0) = u, \text{IIIa}) = \int_{-\infty}^0 \int_0^{L_0} \int_0^{L_0} \sum_{k \in \mathbf{N}} \sum_{j \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\mathbf{y} \in \mathbf{Y}_j^{(-)}} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}(\mathbf{u})} \rho_i(\tau_x, \tau_y, \tau_z, \mathbf{u}) \pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) d\tau_z d\tau_y d\tau_x, \quad (44)$$

where

$$\rho_i(\tau_x, \tau_y, \tau_z, \mathbf{u}) = \begin{cases} 1 & ; \quad \Delta C_i(R_i, t_1, t_2, \mathbf{u}) \leq \theta_i \\ 0 & ; \quad \text{otherwise,} \end{cases} \quad (45)$$

$t_1 = (\tau_x + L_0 - \tau_z)^+$, $t_2 = \tau_y + L_0 - \tau_z$, and π_{kji} is obtained from (35).

Analogously, for Event IIIb we have $\mathbf{V}(\tau_y^{(-)}) = \tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}(\mathbf{u})$,

$$\tilde{\mathbf{Y}}_i^{(-)}(\mathbf{u}) = \{\tilde{\mathbf{y}} \in \mathbb{Z}^{N+1} \mid \tilde{y}_0 = u; \tilde{y}_i = R_i + 1; \tilde{y}_n \leq R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N} \setminus i\}, \quad (46)$$

for $u = R_i + q, R_i + 2q, \dots, Q_i - q$ and $\tilde{\mathbf{Y}}_i^{(-)}(\mathbf{u}) = \emptyset$ otherwise. This results in

$$P(O_i(L_0) = u, \text{IIIb}) = \int_{-\infty}^0 \int_0^{L_0} \sum_{k \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}(\mathbf{u})} \tilde{\rho}_i(\tau_x, \tau_y, \mathbf{u}) \tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) d\tau_y d\tau_x, \quad (47)$$

where

$$\tilde{\rho}_i(\tau_x, \tau_y, \mathbf{u}) = \begin{cases} 1 & ; \quad \Delta C_i(R_i, t_1, t_2, \mathbf{u}) \leq \theta_i \\ 0 & ; \quad \text{otherwise,} \end{cases} \quad (48)$$

$t_1 = (\tau_x + L_0 - \tau_z)^+$, $t_2 = L_0$, 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

$$S_i = \int_{\varepsilon}^{L_0} \frac{f(T)}{T} dT, \quad (49)$$

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

$$\begin{aligned} f(T) &= \frac{q}{\prod_{j=0}^N Q_j} \sum_{\mathbf{a} \in \mathbf{A}} \\ &\int_{-\infty}^0 \int_0^{L_0} \int_0^{L_0} \sum_{k \in \mathbf{N}} \sum_{j \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\mathbf{y} \in \mathbf{Y}_j^{(-)}} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} \sigma_i(\tau_x, \tau_y, \tau_z, \mathbf{z}, T) \pi_{kji}(\tau_x, \tau_y, \tau_z, \mathbf{x}, \mathbf{y}, \mathbf{z}) d\tau_z d\tau_y d\tau_x + \\ &\int_{-\infty}^0 \int_0^{L_0} \sum_{k \in \mathbf{N}} \sum_{\mathbf{x} \in \mathbf{X}_k^{(+)}} \sum_{\tilde{\mathbf{y}} \in \tilde{\mathbf{Y}}_i^{(-)}} \tilde{\sigma}_i(\tau_x, \tau_y, \tilde{\mathbf{y}}, T) \tilde{\pi}_{ki}(\tau_x, \tau_y, \mathbf{x}, \tilde{\mathbf{y}}) d\tau_y d\tau_x, \end{aligned} \quad (50)$$

where

$$\sigma_i(\tau_x, \tau_y, \tau_z, \mathbf{z}, T) = \begin{cases} 1 & ; u < Q_i, \quad t_2 - t_1 = T \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) > \theta_i \\ 0 & ; \text{otherwise,} \end{cases} \quad (51)$$

$u = z_0$, $t_1 = (\tau_x + L_0 - \tau_z)^+$, $t_2 = \tau_y + L_0 - \tau_z$, and

$$\tilde{\sigma}_i(\tau_x, \tau_y, \tilde{\mathbf{y}}, T) = \begin{cases} 1 & ; u < Q_i, \quad t_2 - t_1 = T \quad \text{and} \quad \Delta C_i(R_i, t_1, t_2, u) > \theta_i \\ 0 & ; \text{otherwise,} \end{cases} \quad (52)$$

$u = y_0$, $t_1 = (\tau_x + L_0 - \tau_y)^+$, $t_2 = L_0$. 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

$$\begin{aligned} \Omega(m) &= \{\boldsymbol{\omega} \in \mathbb{Z}^{N+1} \mid \omega_0 = m; \\ &\quad \omega_n = R_n + 1, R_n + 2, \dots, R_n + Q_n, \forall n \in \mathbf{N}\}. \end{aligned} \quad (53)$$

The probability for this is obtained as

$$P(\text{IL}_0(L_0) = m) = \frac{q}{N} \sum_{j=0}^N \sum_{\mathbf{a} \in \mathbf{A}} \sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}(m)} P(\mathbf{V}(L_0) = \boldsymbol{\omega} | \mathbf{V}(0) = \mathbf{a}). \quad (54)$$

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

$$P(\text{IL}_i(L_0 + L_i) = m, \text{III}) = \int_0^{L_0} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} P(D_i(\tau_z, L_0 + L_i) = R_i + u - m) P(\mathbf{V}(\tau_z) = \mathbf{z} | \mathbf{V}(0) = \mathbf{a}) \lambda_i d\tau_z, \quad (55)$$

where $u = z_0$. For the CD policy we have

$$P(\text{IL}_i(L_0 + L_i) = m, \text{III}) = \int_0^{L_0} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}} p_i^{\text{CD}}(\tau_z, \mathbf{z}) P(\mathbf{V}(\tau_z) = \mathbf{z} | \mathbf{V}(0) = \mathbf{a}) \lambda_i d\tau_z, \quad (56)$$

where

$$p_i^{\text{CD}}(\tau_z, \mathbf{z}) = \begin{cases} P(D_i(\tau_z, L_0 + L_i) = R_i + Q_i - m) & ; \quad z_0 = Q_i \\ P(D_i(\tau_z, L_0 + L_i) = R_i - m) & ; \quad z_0 < Q_i. \end{cases} \quad (57)$$

The distribution of IL_0 is independent of the delivery policy and is obtained from (54). The expected number of units on hold, $E[\text{O}_i]$, 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

$$P(\text{O}_i(L_0) = u) = \frac{q}{N} \sum_{j=0}^N \sum_{\mathbf{a} \in \mathbf{A}} \int_0^{L_0} \sum_{\mathbf{z} \in \mathbf{Z}_i^{(-)}(u)} P(\mathbf{V}(\tau_z) = \mathbf{z} | \mathbf{V}(0) = \mathbf{a}) \lambda_i d\tau_z. \quad (58)$$

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 $q - nQ_0 \leq \text{IL}_0(L_0) \leq Q_i - q - nQ_0$ ($n = 0, 1, 2, \dots$). We therefore have

$$S_i = \frac{q}{N} \sum_{j=0}^N \sum_{\mathbf{a} \in \mathbf{A}} \sum_{n=0}^{\infty} \sum_{\boldsymbol{\omega} \in \hat{\boldsymbol{\Omega}}_i(n)} P(\mathbf{V}(L_0) = \boldsymbol{\omega} | \mathbf{V}(0) = \mathbf{a}) \lambda_i, \quad (59)$$

where

$$\hat{\boldsymbol{\Omega}}_i(n) = \{\boldsymbol{\omega} \in \mathbb{Z}^{N+1} \mid \omega_0 = -nQ_0 + q, -nQ_0 + 2q, \dots, -nQ_0 + Q_i - q; \omega_i = R_i + 1; \omega_k = R_k + 1, R_k + 2, \dots, R_k + Q_k, \forall k \in \mathbf{N} \setminus i\}. \quad (60)$$

Optimization of reorder points

Let $TC^{\text{MSD}}(\mathbf{R})$, $TC^{\text{PD}}(\mathbf{R})$ and $TC^{\text{CD}}(\mathbf{R})$ be the total expected cost for each respective delivery policy, given a set of reorder points $\mathbf{R} = (R_0, R_1, \dots, R_N)$. Furthermore, let \mathbf{R}^{MSD} , \mathbf{R}^{PD} 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 $\theta_1 = \theta_2 = \dots = \theta_N = 0$. This special case has been analyzed previously (e.g. Axsäter, 2000) and we refer to it as the PD0 policy, with expected cost $TC^{\text{PD0}}(\mathbf{R})$.

We wish to minimize

$$TC^*(\mathbf{R}) = h_0(E[IL_0^+]) + \sum_{i=1}^N E[O_i] + \sum_{i=1}^N \theta_i S_i + \sum_{i=1}^N (h_i E[IL_i^+] + b_i E[IL_i^-]). \quad (61)$$

We know that for a given R_0 , $TC^{\text{PD0}}(\mathbf{R})$ is separable and convex in the retailer reorder points. It is obvious that the same holds true for $TC^{\text{PD}}(\mathbf{R})$. 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 ($i = 1, 2, \dots, N$). 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 , $TC^{\text{CD}}(\mathbf{R})$ is separable and convex in R_i , $i = 1, 2, \dots, N$.

Proof

As stated previously, $TC^{\text{CD}}(\mathbf{R})$ 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 $TC^{\text{MSD}}(\mathbf{R})$ 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 $R_i^*(R_0)$ denote the optimal reorder point given R_0 for a specific delivery policy (in case the optimal solution is not unique, let $R_i^{\text{PD}}(R_0)$ be the smallest - and $R_i^{\text{CD}}(R_0)$ 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 $i = 1, 2, \dots, N$.

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 ($j = 1, 2, \dots, N$). 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

$$TC^{MSD}(\mathbf{R}) = TC^{PD}(\mathbf{R}) - E_{u, t_1, t_2} \left[\sum_{j=1}^N \mu_j(u, t_1, t_2) (\theta_j - \Delta C_j(R_j, t_1, t_2, u), 0)^+ \right]. \quad (62)$$

Now assume that $R_i^{MSD}(R_0) < R_i^{PD}(R_0)$. This implies that

$$TC^{PD}(R_0, R_1, \dots, R_i^{MSD}(R_0), \dots, R_N) > TC^{PD}(R_0, R_1, \dots, R_i^{PD}(R_0), \dots, R_N). \quad (63)$$

Furthermore, if $\Delta C_i(R_i, t_1, t_2, u)$ is decreasing in R_i ,

$$(\theta_i - \Delta C_i(R_i^{MSD}, t_1, t_2, u), 0)^+ \geq (\theta_i - \Delta C_i(R_i^{PD}, t_1, t_2, u), 0)^+, \quad (64)$$

and therefore from (62) we have that

$$TC^{MSD}(R_0, R_1, \dots, R_i^{MSD}(R_0), \dots, R_N) > TC^{MSD}(R_0, R_1, \dots, R_i^{PD}(R_0), \dots, R_N). \quad (65)$$

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

$$TC^{MSD}(\mathbf{R}) = TC^{CD}(\mathbf{R}) - E_{u, t_1, t_2} \left[\sum_{j=1}^N \mu_j(u, t_1, t_2) (\Delta C_j(R_j, t_1, t_2, u) - \theta_j, 0)^+ \right], \quad (66)$$

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

$$\Delta C(R_i + 1, t_1, t_2, u) - \Delta C(R_i, t_1, t_2, u) = \sum_{n=1}^u \Delta g(R_i, n), \text{ where} \\ \Delta g(R_i, n) = \begin{cases} \frac{h_i + b_i}{\lambda_i} \left(\sum_{j=0}^{R_i+n} \frac{(\lambda_i(t_2 + L_i))^j}{j!} e^{-\lambda_i(t_2 + L_i)} - \sum_{j=0}^{R_i+n} \frac{(\lambda_i(t_1 + L_i))^j}{j!} e^{-\lambda_i(t_1 + L_i)} \right); & R_i + n \geq 0 \\ 0 & ; R_i + n < 0. \end{cases} \quad (67)$$

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 $R_i^*(R_0)$, $i = 1, 2, \dots, N$. (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 $h_0IL_0^+(R_0)$ 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.

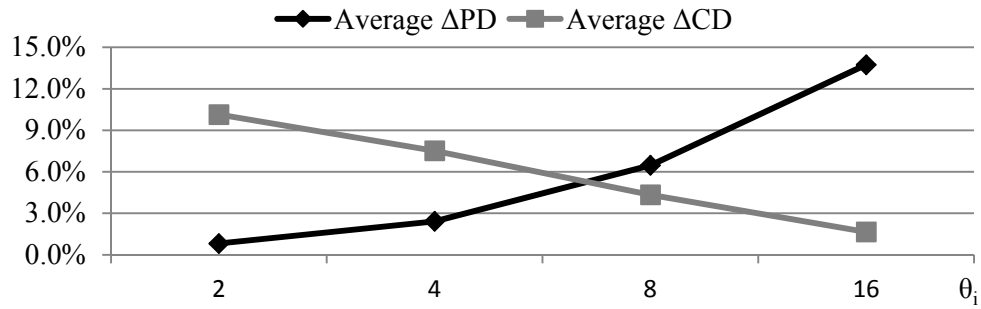


Figure 5

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).

Table 1.

#	Input data				Results MSD		Results PD		Results CD		Comparison	
	θ_i	b_i	L_i	Q	R^{MSD}	TC^{MSD}	R^{PD}	TC^{PD}	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%

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).

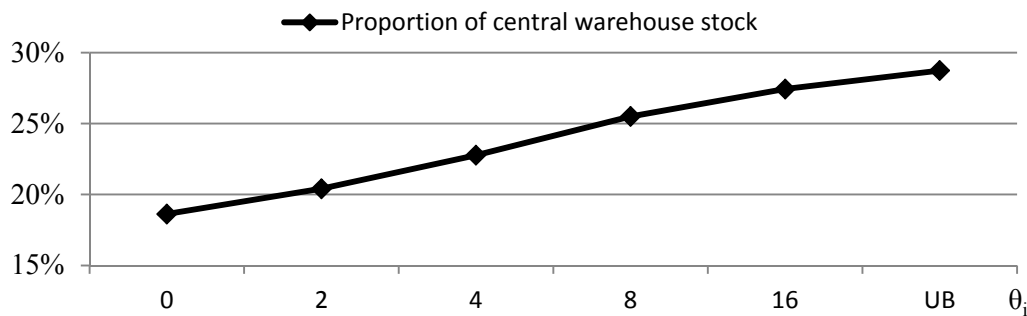


Figure 6

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.