LUND UNIVERSITY

Real-time Allocation Decisions in Multi-echelon Inventory Control

Howard, Christian

2013

Citation for published version (APA):
Howard, C. (2013). *Real-time Allocation Decisions in Multi-echelon Inventory Control*. Lund University (Media-Tryck).

LUND UNIVERSITY

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

Real-time Allocation Decisions in Multi-echelon Inventory Control

LUND
UNIVERSITY

Christian Howard

CLIMATE
COMPENSATED
PAPER

Lund University, Faculty of Engineering LTH
Department of Industrial Management and Logistics
Division of Production Management
ISBN 978-91-7473-509-3

Printed in Sweden by Media-Tryck, Lund University
Lund 2013

To Liselotte, Elicia and Tienne

Abstract

Inventory control is a crucial activity for many companies. Given the recent advances in information technology, there have never been greater opportunities for coordinated inventory control across supply chain facilities. But how do we design efficient control methods and policies that take advantage of the detailed information that is now becoming available? This doctoral thesis investigates these issues within the field of inventory control theory. The objective of the research is:

To develop mathematical models and policies for efficient control and increased understanding of stochastic multi-echelon inventory systems, with a focus on allocation decisions and the use of real-time information.

This thesis is based on five scientific papers which are preceded by a summarizing introduction. The papers address different types of inventory distribution systems, all consisting of a central stocking facility that supplies an arbitrary number of local stocking facilities (referred to as retailers). The retailers face stochastic end customer demand. The systems are characterized by the presence of real-time inventory information, including continuously updated information on the current inventory levels at different facilities and on the locations of outstanding orders.

In Paper I and Paper II we derive and evaluate different decision rules for stock allocation (known as allocation policies) for a central warehouse which applies a time based shipment consolidation strategy. The allocation policy determines how the central warehouse should distribute its stock among different retailers in case of shortages. New allocation policies that utilize real-time information are compared to the commonly used First Come - First Served policy which requires less information.

In Paper III we shift focus to the delivery policy at a central warehouse which supplies multiple retailers that order in batches. When the central warehouse cannot satisfy an entire retailer order immediately, the delivery policy determines if the order should be shipped in several parts or in its entirety when all items are available. We investigate the value of using a new delivery policy that uses real-time information on when replenishments will arrive at the central warehouse. The information is used to determine the best course of action for each order placed by the retailers. We also study how to allocate safety stocks to all facilities in the system given this new policy.

In Paper IV and Paper V we consider a system where retailers may receive emergency shipments from a support warehouse in combination with regular replenishments from a central warehouse/outside supplier. We investigate how safety stocks should be allocated between the retailers and the support warehouse. Furthermore, we evaluate the benefits of tracking orders in real time and using this information in the decision whether or not to request an emergency shipment.

Keywords: Inventory, Multi-echelon, Real-time information, Stochastic, Stock allocation

Acknowledgements

Looking back at the journey that has led to this doctoral thesis, there are undoubtedly many people that have helped me reach this stage. First of all, I would like to thank my main supervisor Johan Marklund. Johan has guided me through the PhD process, providing me with the feedback and help I have needed. He is one of those rare people who manages to combine ambitious goals and requirements with an open and compassionate attitude, and I am very grateful to have had him as my main supervisor.

I would also like to thank my assistant supervisor Sven Axsäter. In addition to working together on one of the papers in this thesis, Sven’s door has always been open when I have needed help. He has on countless occasions taken the time to answer my questions, read my unfinished work and provide me with suggestions for improvement.

Thanks to Olle Stenius who has been my brother-in-arms and friend throughout the entire PhD journey. Olle has been with me through tough revision processes, nervous presentations, complex derivations, and thrilling encounters with exotic animals (and relatives) in Australia. His contribution to the completion of this thesis cannot be overstated.

To the “inventory people” at the division: Peter Berling, Lina Johansson and Fredrik Olsson: Thank you for your friendship, help and support. This thesis has benefited greatly from your knowledge. A special thanks to Peter for not laughing (too loud) at my sometimes trivial and stupid questions. I would also like to thank Tarkan Tan and Ingrid Reijnen at Eindhoven University of Technology for the fruitful collaboration that resulted in the fifth paper of this thesis. Furthermore, thanks to Christian Larsen at Aarhus University for being the discussant at my licentiate seminar.

Thanks to all my friends and colleagues at the university, past and present. To Fredrik Eng-Larsson for sharing my weird and vicious sense of humor. To Ala Pazirandeh for all the lasagna, gingerbread and “gold corns”. To Ali Pazirandeh for teaching me that there are always two sides to an argument. To Kostas Selviaridis for enduring an endless stream of jokes related to the state of the European economy, and to Hoda Davarzani, Hana Hulthén, Carina Johnsson, Joakim Kembro, Johan Lundin, Fredrik Nilsson, Birgitta Olvegart, Sebastian Pashaei, Patrik Tydesjö, Christina Öberg and everyone else at the department.

To my oldest and dearest friends, Matteo Beghello and Sebastian Jacobsson: Thank you for being there through thick and thin.

Last, but certainly not least, I thank my family: my partner Liselotte, and our wonderful daughters Elicia and Tienne, for putting up with my highs and my lows and for your endless love and support; my parents Richard and Lilian, and my sister Anna, for always being there and always believing in me.

Now, there are most likely additional people who deserve an acknowledgement. I apologize for forgetting to mention you and I blame my scattered researcher mind. I promise that you will be acknowledged in my next doctoral thesis…

Contents

1. List of papers

- **Paper I – Evaluation of Stock Allocation Policies in a Divergent Inventory System with Shipment Consolidation**
  Howard, C., J. Marklund
  (2011) *European Journal of Operational Research* **211** 298-309
- **Paper II – New Allocation Policies for Divergent Inventory Systems with Real-time Information and Shipment Consolidation**
  Howard, C.
  Department of Industrial Management and Logistics, Lund University
  (2013) Under review
- **Paper III – Partial or Complete Deliveries in Two-echelon Inventory Systems?**
  Howard, C., O. Stenius
  Department of Industrial Management and Logistics, Lund University
  (2013) Under review
- **Paper IV – A Distribution Inventory Model with Transshipments from a Support Warehouse**
  Axsäter, S., C. Howard, J. Marklund
  (2013) *IIE Transactions* **45** 309-322
- **Paper V – Using Pipeline Information in a Multi-echelon Spare Parts Inventory System**
  Howard, C., I. Reijnen, J. Marklund, T. Tan
  Department of Industrial Management and Logistics, Lund University, and
  Department of Industrial Engineering and Innovation Sciences,
  Eindhoven University of Technology
  (2013) Under review

2. Introduction

The past decades have witnessed unprecedented advances in information technology, fundamentally transforming the environment in which businesses operate. The costs for obtaining detailed information on supply chain operations have decreased, and the technological barriers that once prevented the sharing of information across supply chain facilities are steadily dissipating. Therefore, there have never been greater opportunities for efficient control of inventory distribution systems.

The control of inventories is crucial to many companies because profitability is often linked to providing high availability of physical products. One of the main purposes with keeping inventory is to ensure this availability by guarding against uncertainties in demand. However, there are also considerable costs associated with keeping materials and products in stock (cost of capital tied up in inventories, costs for storage space, costs for obsolescence and pilferage etc.). Thus, there is a need for decision rules and policies that balance the allocation of stock with the customer's service requirements. A key question is then how to take advantage of the richer information structures available today when designing these policies. Furthermore, what is the actual value, in terms of cost efficiency and customer satisfaction, of incorporating more detailed information in the decision making? The research that is presented in this thesis investigates these issues within the field of inventory control theory.

This thesis is based on five scientific papers that focus on different two-level distribution systems. The common structure of all inventory systems considered is that they consist of an arbitrary number of retailers (local stock points that are also known as dealers or local warehouses) that face uncertain customer demand. The retailers are replenished by a central stock point. Two different types of central stock points are considered in this work: the first type is a central warehouse that handles all orders placed by the retailers, the second type is an emergency supplier (referred to as a support warehouse) which provides quick deliveries at an extra cost when the retailers require it. The systems considered are characterized by the accessibility to real-time inventory information. In this context this means that the decision maker has access to continuously updated information on current stock levels, and also has access to information on the location of particular orders in the system. Given this availability of detailed inventory information, new methods for allocating safety stocks and new policies for allocating stock from the central stock point to the retailers are developed. Furthermore, evaluating the performance of these new methods and policies provides insights on the value that real-time information can bring to inventory systems.

The disposition of this thesis is as follows: Chapter 1 provides a list of the scientific papers on which this thesis is based. Chapter 2 has so far given a short introduction to the general problem area considered. This chapter will continue with the research objectives, followed by a discussion about the research methodology used. The chapter concludes with a brief introduction to inventory control theory. Chapter 3 provides extended summaries of the appended papers. Chapter 4 concludes with a discussion on the research contributions and an outlook on possible future research directions.

2.1 Research objectives

The objectives of the research presented in this thesis are summarized as:

To develop mathematical models and policies for efficient control and increased understanding of stochastic multi-echelon inventory systems, with a focus on allocation decisions and the use of real-time information.

In this context, the term “efficient control” refers to finding a control method that is optimal (or near optimal) given a certain objective, and given a set of rules for the system. Strictly speaking, in this work the objective is to minimize the costs associated with maintaining an inventory system. The given rules are, for instance, the system structure, information availability, cost structures, service requirements and the stock replenishment policies of the system. The “increased understanding” refers to the insights gained from studying the model’s underlying mathematics and the results obtained by numerically experimenting with different scenarios.

The “allocation decisions” are the decisions regarding *where*, and in *what quantity*, to stock a particular item. The decisions also concern *when* and *how* to allocate an item to a given stock point. As mentioned in the previous section, “Real-time information” refers to an information structure where the decision maker has access to detailed information on the system’s current state. In this work, this predominantly means continuously updated information on current stock levels and on the arrival times of incoming orders.

The five papers on which this thesis is based adhere to the general objective as follows:

In Paper I and Paper II we derive and evaluate different decision rules for stock allocation (known as allocation policies) for a central warehouse which applies a time based shipment consolidation strategy. The allocation policy determines how the central warehouse should distribute its stock among multiple retailers in case of shortages. New allocation policies that utilize real-time information are compared to the commonly used First Come - First Served policy which requires less information.

In Paper III we focus on the delivery policy at a central warehouse which supplies multiple retailers that order in batches. When the central warehouse cannot satisfy an entire retailer order immediately, the delivery policy determines if the order should be shipped in several parts or in its entirety when all items that are needed are available. We investigate the value of using a new delivery policy that uses real-time information on when replenishment will occur at the central warehouse. The information is used to determine the best course of action for each order placed by the retailers. We also study how to allocate safety stocks to all facilities in the system given this new policy.

In Paper IV and Paper V we consider a system where retailers receive emergency shipments, when needed, from a support warehouse. We investigate how safety stocks should be allocated between the retailers and the support warehouse. Furthermore, we evaluate the benefits of tracking orders in real time and using this information in the decision whether or not to request an emergency shipment.

2.2 Research methodology

Inventory control is a subdiscipline of the broader field of operations research and the management sciences. This broader field is characterized by the use of mathematical models to solve complex operational and managerial problems that stem from real life. The solutions are often obtained by using methods from many different areas. For inventory control models, this usually includes areas such as probability theory, queuing theory, control theory, statistical inference, mathematical optimization, computer science and programming, as well as logistics, economics and business administration.

It should be noted that the research objective is “To develop mathematical models…”. Thus, the choice of research method is determined by the objective. However, under a more general objective, for instance, to determine how real-time information can be used to improve supply chain operations, there would be a multitude of feasible alternative methods. For example, one could carry out a case study: using surveys or conducting interviews to determine suitable strategies. The main strength of mathematical modeling is that this method can provide a high degree of generality and objectivity regarding the results obtained. Moreover, there are obviously situations regarding data analysis and parameter determination where using any other method would not be feasible. However, a weakness of mathematical modeling is that the models can be sensitive to particular assumptions made. Therefore they may not provide a relevant depiction of reality. Discussions on the relevance of the models considered in this thesis are provided in the individual papers.

Generally speaking, building a mathematical model is an iterative process that usually contains three separate steps (e.g., Hillier and Lieberman, 2010, and Axsäter and Marklund, 2010).

The first step is the formulation of a model. That is, one seeks to find a set of rules that adequately describes the problem studied. The motivation for studying the problem may be general in nature, stemming from broad observations of a common problem. The motivation can also be more direct and come from a problem observed in a specific organization. In this work, Paper I, Paper II and Paper III fit the former description, where rather general problems that occur in many different types of inventory systems are studied. Conversely, Paper IV and Paper V fit the latter description, where the models considered stem from the specific needs of a large spare parts service provider. Obviously, although these models are grounded on the conditions at this service provider, the aim is to provide results and insights that are relevant for other types of systems as well.

When formulating a mathematical model, one cannot hope for the model to capture all real life features in their entirety. On the contrary, one of the main challenges is to find the right level of detail. A model that is too complex makes it impossible to extract any useful results, while an over-simplified model does not tell us anything about the real problem studied.

The second step concerns the analysis of the model formulated. That is, a method for evaluating the model’s behavior is derived. In this work, this means developing analytical methods for obtaining performance measures and optimizing the inventory system studied. As is common when studying complex systems, the analysis usually incorporates some approximations. Discrete-event simulation is in these cases used to validate the approximations, and to obtain any performance measures that cannot be obtained analytically. Simulation is used to a varying extent in Paper I, Paper II, Paper IV and Paper V. However, it is important to point out that simulation is merely used as a tool for evaluation of the analytical models. The aim is to develop analytical – not simulation based – methods for inventory control.

In the third step one validates that the model gives an accurate description of the real system studied. Depending on the results found, one might have to return to previous steps and review the assumptions made. The number of iterations needed, before a final model is obtained, usually depends on the complexity of the problem studied.

The use of the final model depends on the initial purpose of analyzing the real problem. In inventory theory, a model is often designed to be a concrete tool for finding solutions to problems associated with operating and controlling the system. For instance, a model may be used to determine how many units to keep in stock at a specific location. Ideally a mathematical model will also provide managerial insights and promote a general understanding of the behavior of the system considered. As is the case in this thesis, many insights are also gained from numerical experimentation. That is, by studying how the model reacts to changes in input data and drawing conclusions thereof.

2.3 Inventory control theory

In this section some key concepts within inventory control are presented. The purpose is to show how this current work relates to previous work within the field on a general level. More detailed discussions about previous work and how the individual papers are positioned in the literature are provided in each of the appended papers. The unfamiliar reader who wishes to learn more about inventory control can also turn to, for example, Silver et. al. (1998), Zipkin (2000) and Axsäter (2006).

2.3.1 Single-echelon systems

The most basic feature of any inventory system is its structure (also known as topology). The structure tells us the number of stock points that are included in the system, and how these stock points are connected to each other. The simplest structure is the single-echelon system, depicted in Figure 1.

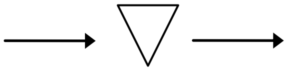

Figure 1. The single-echelon system

In this system, a single stock point faces demand that depletes the inventory on hand, while replenishments arrive from a supplier. The supplier is outside the scope of the system and delivers according to a lead time that is either constant or stochastic. The stock point can hold a single type of item, or many different items. Although a real system most likely holds many different kinds of items, it is often the case that these can be controlled separately, implying that a single-item model often is sufficient.

An important feature is the demand process. This thesis is focused on systems with stochastic demand. For these types of systems, a common and often reasonable assumption is that demand follows a Poisson process, a compound Poisson process, or a more general renewal process. The use of the normal distribution as an approximation of discrete demand is also widespread.

The assumptions regarding the cost structure is another key issue. Typically one assumes an inventory holding cost for items physically in stock, to be balanced with a service constraint or a penalty cost for providing poor service. The form of the latter two is related to what happens when a stock-out occurs: In some systems demand is backordered, in other it is lost. A fixed cost for placing an order independent of its size and a variable volume dependent cost are also common in many inventory systems.

Information availability is a central theme in this work. Typically, systems are categorized based on how often the status of the system is inspected. When the status is monitored continuously the system is referred to as a continuous review system, otherwise it is known as a periodic review system. Furthermore, one needs to define exactly what information is available when the system is inspected. It is common that the information is limited to knowing the inventory position (stock on-hand + outstanding orders – backorders) at the stock point considered. However, following the advances in information technology in practice, systems with extended information structures (like the ones considered in this work) are becoming more common.

Given the characteristics that define the system, the objective is to find the optimal control policy. That is, to determine when to place replenishment orders, and the sizes that these orders should be. For the single-echelon system 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). This control policy implies that an order to bring the inventory position to S is placed when the inventory position is observed to reach or drop below s units. For an efficient method for determining the optimal values of s and S, see Zheng and Federgruen (1991).

The papers mentioned above assume that there is only one type of customer in the system. That is, that all customer demands have equal importance. There are also a number of papers that consider multiple customer demand classes (see, e.g., Kleijn and Dekker, 1998, Arslan et al., 2007, Teunter and Haneveld, 2008). The demand classes have different importance and this is translated into different backorder costs or service requirements. As a result, the challenge is not only to determine the best ordering policy, but also how to ration stock among the different demand classes. A common approach is to assume a so-called critical-level policy, where only demands with higher importance are satisfied when the inventory reaches a certain critical level.

For many systems it might be very difficult to derive the optimal control policy, especially if the structure is more complicated than in the single-echelon system. Furthermore, the form of the optimal policy can sometimes be very complex, making it impractical for implementation in practice. It is therefore common to assume a relatively simple control policy, and focus on optimizing the parameters of that given policy. A widely used control policy, both in theory and in practice, is the (R,Q) policy. It is similar to the (s,S) policy, except that it involves fixed order quantities. To be precise, a batch of Q units is ordered when the inventory position reaches or drops below the reorder point R. Another well-known policy is the base-stock policy. It implies that an order is placed to raise the inventory position up to the base-stock level S each time the inventory status is reviewed (it is also known as an order-up-to-S policy). Hence, one will always order an amount equivalent to the demand taken place from the time of the last review. The policy is a special case of the (s,S) policy, with $s = S-1$, and in a continuous review system it means that items are ordered concurrently as demand occurs.

For a more comprehensive overview of models concerning single-echelon systems, see for example Lee and Nahmias (1993). The classic work by Hadley and Whitin (1963) also provides a good background to the ideas and models discussed in this section.

2.3.2 Multi-echelon systems

Multi-echelon systems feature multiple stock points that are connected to each other through the replenishment structure. Therefore, they are generally more difficult to analyze than single-echelon systems. The simplest type of multi-echelon structure is the serial system, where each stock point only has one single immediate predecessor and one single immediate successor. Figure 2 shows an example of a three-echelon serial system.

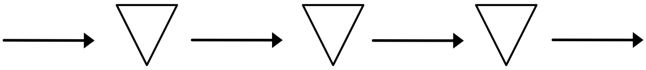

Figure 2. A serial system

There exist some optimality results for serial systems. In the case of ordering costs only at the most upstream facility, it optimal to use an (s,S) policy at this facility and (S-1,S) policies at remaining downstream facilities (Clark and Scarf, 1960, Federgruen and Zipkin, 1984). Under the assumption that items flow in fixed batches Chen (2000) shows that (R,nQ) policies are optimal. Chen (1999), Muharremoglu and Tsitsiklis (2008) provide overviews and further results on serial systems.

The assembly system (exemplified in Figure 3) is common in many production processes, where different components are put together to form a final product. Its main characteristic is that each stock point only has one immediate successor.

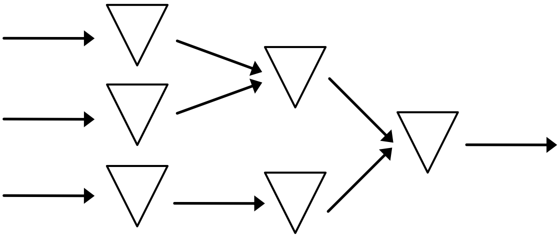

Figure 3. An assembly system

Rosling (1989) shows that, under certain (rather weak) conditions, assembly systems can be decomposed into a number of serial systems. Hence, many of the results for serial systems carry over to the assembly system, including the optimality results by Clark and Scarf (1960) and Federgruen and Zipkin (1984).

The divergent multi-echelon distribution system is another common structure. It is also the focal point of Paper I, Paper II and Paper III. The system's distinguishing feature is that each stock point only has one immediate predecessor, but may have multiple immediate successors. This makes it more general than the serial system. An example of a two-echelon distribution system with one central warehouse and three retailers is depicted in Figure 4.

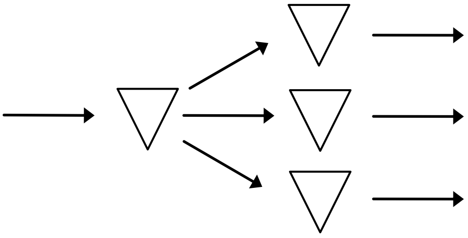

Figure 4. A divergent distribution system

A question that arises in these types of systems, which is also the topic of Paper I and Paper II, is what kind of allocation policies to use at the upper echelons. That is, how should the central warehouse distribute its stock among the downstream facilities? For example, if the central warehouse does not have sufficient stock on hand, in what sequence should orders from the retailers be satisfied? In the literature on periodic review systems it is common to use so-called myopic allocation policies (first presented by Clark and Scarf, 1960). These policies essentially imply that an optimization problem is solved each time items are dispatched from the central warehouse. In continuous review systems, much of the analysis is based on the simple First Come - First Served (FCFS) policy (e.g., Axsäter, 2003, 2007). This policy implies that orders are satisfied in the sequence that they occur (the policy is also present in the periodic review literature, e.g., Axsäter 1993, Graves, 1996, Shang and Zhou, 2011). The popularity and simplicity of the FCFS policy makes it interesting to evaluate how well it works in terms of cost performance. Thus, in Paper I the performance of the FCFS policy is evaluated by comparing it to policies that utilize a richer information structure. This is done by applying different myopic allocation policies in a two-echelon system with real-time information and a shipment consolidation policy at the central warehouse. The shipment consolidation policy is time based, meaning that shipments leave the central warehouse to visit all retailers according to a predefined time interval. In Paper II the consolidation policy allows for several retailer groups with different shipment time intervals. Thus, Paper II considers a more general system than Paper I. In Paper II two additional allocation policies are derived. These new policies are based on a different method of analysis and results show that they outperform the FCFS policy, as well as the policies in Paper I.

The issues of allocation, combined with the dependencies between the different stock points, imply that analysis of the divergent distribution system is complex. Therefore, there are currently no optimality results for the general N-echelon system. However, there are some results for special cases of the two-echelon system. For example, under the assumption of (R,Q) policies at all lower echelons, Axsäter and Marklund (2008) derive an ordering policy at the central warehouse that is optimal in the class of position based policies. This class includes all known continuous review policies for which exact evaluation methods are available to date. There is also a large body of literature for optimizing the parameters of various given replenishment and allocation policies. For more information on this literature see, for example, Axsäter (2003), Axsäter and Marklund (2008), Chu and Shen (2010) and references therein.

There is one issue that is often overlooked in the literature on exact evaluation of divergent two-echelon systems under continuous review. That is the issue of what delivery policy to use at the central warehouse. Assume that a retailer places an order and that the central warehouse can only satisfy part of this order immediately. In the previous literature it is predominantly assumed that the part available is delivered instantly, and the rest is delivered at a later time. However, it might be more cost efficient to wait for replenishment and deliver the entire order at the same time. Paper III investigates these issues by comparing different types of delivery policies. One of the main contributions of this paper is the exact analysis of a delivery policy that makes an optimal delivery decision between partial or complete delivery for each order placed by the retailers.

2.3.3 Multi-supplier systems

All of the models mentioned so far assume that there is only one supplier to choose from when placing a replenishment orders. However, in practice multi-supplier sourcing is quite common. This is also reflected in the literature, where models with multiple modes of supply are analyzed frequently. However, the complexities of these models generally rule out the possibility of deriving optimal policies.

Figure 5 depicts a single-echelon system where a stock point can choose between two different suppliers. A typical assumption for this kind of system is that one supplier offers cheaper replenishments, but with longer replenishment lead times. Conversely, the other supplier has shorter lead times, but is more expensive to use. A common policy under these circumstances is to utilize the cheaper supplier for regular replenishment and to order from the more expensive supplier in emergency situations, for example, when a stock-out occurs (e.g., Moinzadeh and Schmidt, 1991, Song and Zipkin, 2009, Veeraraghavan and Scheller-Wolf, 2008).

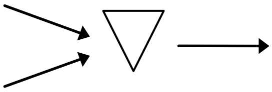

Figure 5. A single-echelon system with two suppliers

As part of the analysis of a larger multi-echelon system, Paper V provides exact results for a new ordering policy, referred to as an (S,T) policy, in a single-echelon system with two suppliers. The policy uses real-time information on when outstanding orders will reach the considered stock point for determining when to use the emergency supplier.

Another form of multiple sourcing is to use lateral transshipments. Lateral transshipment means that a number of stock points in the same echelon share their inventory in some way. Similarly to the single-echelon system above, this usually implies that an inventory location will request a shipment from another stock point when facing a stock-out situation. There is a large variety of different lateral transshipment models described in the literature, ranging from models with only two stock points (Figure 6) to more general multi- echelon structures (exemplified in Figure 7).

Paterson et al. (2011) provides a recent overview of the lateral transshipment literature. For a review of both single and multi-echelon models with multiple suppliers, including lateral transshipment models, see Minner et al. (2003).

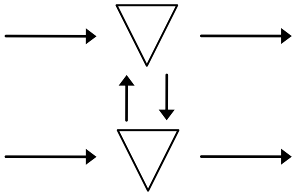

Figure 6. A single-echelon system with lateral transshipments

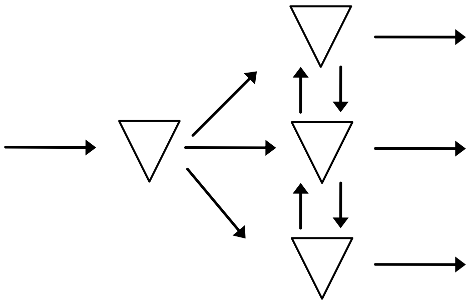

Figure 7. A multi-echelon system with lateral transshipments

Paper IV and Paper V both deal with the issues of dual sourcing, focusing on a system with multiple retailers that can choose between a regular supplier and an emergency supplier (Figure 8).

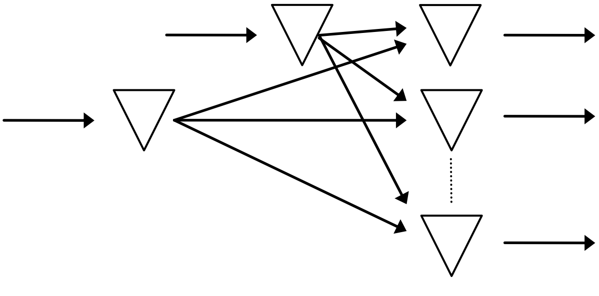

Figure 8. The system considered in Paper IV and Paper V

The retailers face customer demand and place regular replenishment orders to a central warehouse. It is assumed that the central warehouse can always deliver on time and it can hence also be viewed as an outside supplier. In Paper IV emergency orders are placed to a support warehouse when a stock-out occurs. In Paper V emergency orders are placed to the support warehouse or the central warehouse according to the previously mentioned (S,T) policy which utilizes real-time information on incoming orders.

The structure studied in Paper IV and Paper V can be viewed as a special case of a distribution system with lateral transshipments and partial pooling (e.g., Kranenburg and van Houtum, 2009). Partial pooling means that only some stock points can provide transshipments, in our case a single support warehouse that does not face any direct customer demand. On the one hand, this implies that the structure considered in Paper IV and Paper V is less general than in some of the previous models in the literature. On the other hand, the somewhat simpler structure in our work allows for more general assumptions regarding other modeling features, for instance, with respect to ordering policies, replenishment lead times and the use of real-time information.

Another way to view the structure in Paper IV and Paper V is as an extension to the single-echelon system depicted in Figure 5. That is, we add multiple stock points and include the second echelon. In some sense, this is a more intuitive way of viewing our work. This is because our methodology is based on first analyzing a single stock point, and then extending the analysis to include additional facilities.

2.4 Summary of modeling features

The systems considered in the appended papers share several specific features. In accordance with the central themes of this thesis, all systems are stochastic two-echelon continuous review systems. An additional common features is that all transportation times within the systems, as well as all lead times to any outside suppliers, are constant. Furthermore, all papers deal with single-item systems.

Except for the given similarities, the preceding section shows that the research presented in this thesis covers a range of different types of systems and problems. In Table 1 an overview of each paper's key characteristics is provided. The purpose with Table 1 is not to provide a complete summary of each paper (this is provided in Chapter 3), but rather to highlight some key differences between the different papers and to summarize the scope of the presented work.

Table 1. Overview of key characteristics for appended research papers

		Paper				
System features		**I**	**II**	**III**	**IV**	**V**
Structure:	Central warehouse - N retailer	X	X	X		
	Support Warehouse - N retailer				X	X
Real-time information:	Inventory positions	X	X	X	X	X
	Replenishment orders	X	X	X		X
Demand:	Poisson	X	X	X	X	X
	Compound Poisson				X	
	Normal				X	
Replenishment policy upper echelon:	(S-1,S)	X	X	X	X	X
	(R,Q)	X	X	X	X	
	(S,T)					X
Replenishment policy lower echelon:	(S-1,S)	X	X	X	X	X
	(R,Q)			X	X	
	(S,T)					X
Main focus:	Optimize replenishment policies			X	X	X
	Allocation / delivery policy	X	X	X		
	Emergency shipments				X	X

Note that the (S-1,S) policy is a special case of both the (R,Q) policy and the (S,T) policy.

3. Summary of papers

In this chapter summaries of the five scientific papers are provided. For each paper a general overview of the system structure and features, motivation and research objectives, analysis, results and conclusions is given. Note that detailed model features are presented before the actual motivation for studying the system is discussed. This might seem counterintuitive in light of the mathematical modeling steps discussed in Section 2.2. However, the reason for this structure is simply ease of exposition and clarity regarding the systems studied.

3.1 Paper I – Evaluation of Stock Allocation Policies in a Divergent Inventory System with Shipment Consolidation

In Paper I we consider a system with a central warehouse and a number of retailers (Figure 9).

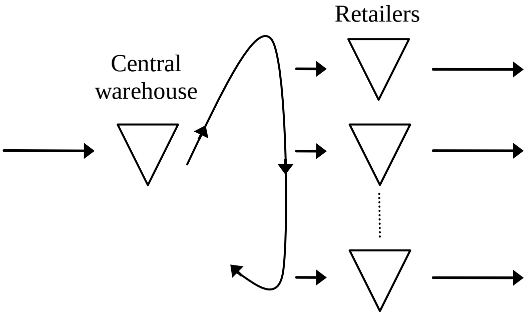

Figure 9. The system considered

The retailers face Poisson customer demand and immediately convey the demand to the central warehouse (i.e., the central warehouse has access to real-time information on point-of-sale data). The central warehouse applies a shipment consolidation policy where the retailers are replenished at fixed time intervals. This means that a shipment (e.g., a single truck) is dispatched from the central warehouse every T time units to visit all retailers sequentially. The warehouse itself orders from an outside supplier using an (R,Q) policy. Unmet demand is backordered at all stock locations. The cost structure includes inventory holding costs per unit and time unit at all stock points, and backorder costs per unit and time unit at the retailers.

The motivation for studying such a system stems from the recent developments in information technology. New tracking and tracing systems increase the availability of real-time demand and inventory information across entire supply chains. Hence, the system captures a situation where point-of-sale data is immediately transferred upstream in the supply chain, from the retailers to the central warehouse (note that from a modeling perspective this is equivalent to the retailers using (S-1,S) policies). At the same time, higher fuel prices and tighter environmental legislation increase the importance of shipment consolidation strategies. Furthermore, despite the access to real-time information there may still be economies of scale associated with ordering from the external supplier (which, for instance, could be a manufacturing facility). This motivates the use of an (R,Q) policy at the central warehouse.

Given the shipment consolidation policy, it follows that the central warehouse needs to decide how much stock to allocate to each of the retailers when a shipment is dispatched. An allocation policy that is easily implemented in practice and commonly used in the literature is the First Come - First Served (FCFS) policy. It implies that orders from the retailers are satisfied in the sequence that they occur, or equivalently for the considered system, in the sequence that customer demand occurs. However, we know that this policy is not optimal and our objective is to benchmark this simple method of allocation against more advanced methods. We consider two alternative state-dependent myopic allocation policies (the term myopic meaning that the allocation decisions are based on a finite time horizon). The first policy is denoted MAs (Myopic Allocation at the moment of shipment) and implies that the allocation decision is made at the moment a shipment is dispatched from the central warehouse. The second policy, MAd (Myopic Allocation at the moment of delivery), postpones the allocation decision until the arrival at the different retailers.

The paper is an extension to the work by Marklund (2011), which analyses the considered system under the assumption of FCFS. The use of myopic allocation policies dates back to the seminal work by Clark and Scarf (1960). They show that myopic allocation is optimal under the idealized assumption that all retailers can redistribute their stock in each time period (often referred to as the balance assumption). Since then, these types of policies have been used extensively in the literature, although mainly in periodic review systems. For papers that specifically evaluate the FCFS allocation policy we mention Graves (1996) and Axsäter (2007). Similar to this current work, Graves considers a system with fixed shipment intervals. However, key differences compared to the present work is that Graves considers a periodic review system, places certain restrictions on the shipment intervals and does not allow for batch ordering at the central warehouse. Moreover, the paper compares FCFS to a lower bound, while we consider feasible alternative allocation policies. Axsäter, on the other hand, analyzes the FCFS policy in a continuous review system. However, two distinct differences are that the central warehouse uses a base-stock policy and that shipments to the retailers are dispatched instantly. The latter excludes the possibility of using an allocation policy at the moment of delivery, but the paper presents a number of heuristics that produce similar results to our MAs policy.

When applying our myopic policies, we base the allocation decision on the expected costs over a single delivery cycle (i.e., the time between two successive deliveries to the retailers). The allocation problem is then formulated as a constrained version of the well known newsvendor problem, and solved by an iterative Lagrangian method. The procedure is essentially identical for MAs and MAd. The only difference is that the MAs policy means that we solve the problem once every time a shipment is dispatched from the central warehouse, while MAd implies that we repeat the procedure upon arrival at each individual retailer. Due to the complexities of the state-dependent policies considered, we use simulation to evaluate the expected costs. Moreover, we use the R and S-values for FCFS allocation which can be obtained using the method in Marklund (2011).

The cost performance of MAs and MAd is investigated in a numerical study. The relative decrease in expected total costs compared to FCFS is used to measure the performance. The study features a basic test series of 192 problem scenarios. Based on the results from the study, we also perform additional tests to complement the basic test series. In the study MAs performed only slightly better than FCFS. The average cost decrease for MAs was 1.6%, with a maximum of 4.3% and a minimum of −2.4%. Note that a negative value means that the policy performed worse than FCFS. The MAd policy performed better, rendering an average cost decrease of 5.6%, with a maximum of 19.5% and a minimum of 0.7%. Furthermore, the study indicates that myopic allocation is most fruitful (i.e., FCFS performs worse) in systems with long transportation times between the central warehouse and the retailers, and in systems with small batch sizes at the central warehouse.

Our study shows that there are situations where an allocation policy based on real-time information provides significant cost savings. However, when also taking simplicity and ease of implementation into consideration, we conclude that FCFS remains an attractive policy to use in many cases.

3.2 Paper II – New Allocation Policies for Divergent Inventory Systems with Real-time Information and Shipment Consolidation

In Paper II we consider a more general version of the system in Paper I. The retailers can now be divided into separate groups (e.g., based on geographical location). Each group of retailers has its own shipment interval, where shipments again leave the central warehouse at fixed time intervals (Figure 10). All other system features are identical to those in Paper I.

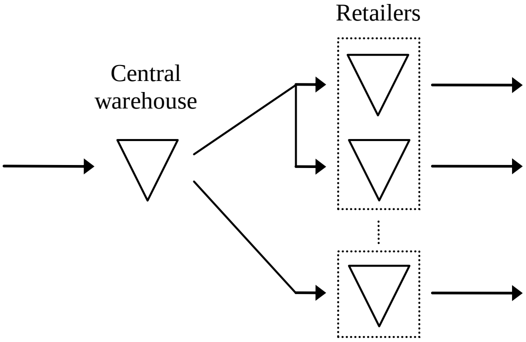

Figure 10. The system considered

The purpose is to evaluate two new allocation policies. The first policy, which is simply referred to as the HAs (Heuristic Allocation at the moment of shipment) policy, is an alternative to the MAs policy. Similarly, the second policy, HAd (Heuristic Allocation at the moment of delivery), is an alternative to the MAd policy. The benefit of our new policies is that they provide a performance guarantee compared to the FCFS policy. That is, our policies will never perform worse than FCFS. As illustrated in the numerical studies in Paper I, no such guarantees can be given for the MAs and the MAd policies. Furthermore, our new policies are derived assuming multiple retailer groups. This allows us to investigate how the number of retailer groups affects the performance of FCFS. Multiple retailer groups with different shipment intervals add additional complexity to the allocation decision. This is because a unit will leave the central warehouse at different times depending on which retailer (and thus retailer group) it is allocated to.

The analysis of our new policies is based on further developing a basic allocation idea in Axsäter (2007). The HAs and HAd policies are derived by studying the system under FCFS allocation. Given FCFS, it is possible to calculate exactly the expected cost that a given unit will incur before it leaves the system. The idea behind the new policies is thus to calculate the expected cost difference of changing retailer assignment between two given units. If a change results in a lower expected cost, a reassignment is performed, providing a guaranteed decrease in total costs (in the long run). Our new policies are based on performing repeated pairwise reassignments for all relevant units each time a shipment leaves the central warehouse (for HAs), or each time a shipment reaches a retailer (for HAd). As in Paper I, we use the optimal R and S values given FCFS which are obtained from the method in Marklund (2011), and we use simulation to evaluate the expected total costs given the new policies.

We perform two numerical studies that together consist of 1,064 problem scenarios. In the first study, which features scenarios with only a single retailer group, the HAs and HAd policies are compared to the MAs and MAd policies. As in Paper I, the relative decrease in expected total costs compared to FCFS is used to measure the performance of each policy. In the study, HAs performed 0.5 percentage points better than the MAs policy, on average. Similarly, the HAd policy performed 0.4 percentage points better than the MAd policy, on average. Although these differences are small, the slightly better average performance combined with the performance guarantee makes it attractive to use our new policies.

In the second study, which features scenarios with up to three different retailer groups, the HAs and HAd policies are compared to the FCFS policy. For the HAs policy, the expected cost decrease compared to FCFS was, on average, 2.4% (maximum 5.5%). The same value for the HAd policy was 6.7% (maximum 19.4%). Our study also indicates that the HAs policy performs better, and the HAd policy performs worse, when a given number of retailers are divided into smaller groups. Another noteworthy result is that both our new policies tend to perform better when the shipment time intervals are short.

Our studies indicate that the results in Paper I carry over to systems with multiple retailer groups. That is, FCFS performs reasonably well in most cases compared to allocation at the moment of shipment, but performs considerably worse compared to allocation at the moment of delivery.

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

In Paper III we consider a central warehouse that supplies a number of retailers (Figure 11). Customer demand follows independent Poisson processes and all stock points use installation stock (R,Q) policies for replenishment. All unmet demand is backordered. Moreover, we consider inventory holding costs per unit and time unit at all stock points, combined with backorder costs per unit and time unit at the retailers.

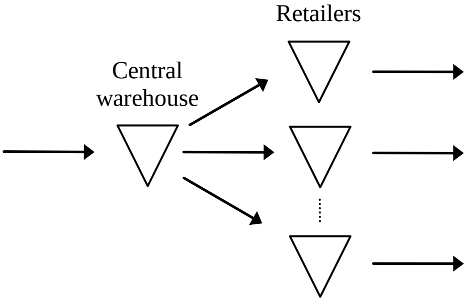

Figure 11. The system considered

In previous exact analysis of the system considered, it is generally assumed that the central warehouse uses a partial delivery policy (Axsäter, 2000, Forsberg, 1997, Marklund, 2002). This means that if a shortage occurs, any units available are shipped to the retailers immediately. Hence, a retailer that orders Q units can receive this order in several parts at different times. However, past models do not consider the extra costs that may occur when an order is split (e.g., costs for repeating activities such as order picking, loading, shipping, unloading, receiving, inspection, authorization and invoicing). The environmental consequences of repeated shipments are also ignored. Thus, an obvious alternative to the partial delivery policy is the complete delivery policy, where units are only shipped in complete batches (to the best of our knowledge Andersson, 1999, provides the only exact solution to this problem). However, in some situations the complete delivery policy will be far from optimal. This would, for instance, be the case when the retailers desperately need replenishment and an incomplete order is held up at the central warehouse.

In this work we introduce a handling cost that quantifies the extra cost for partial delivery, that is, the extra cost for splitting a retailer order. Given this new cost parameter it is possible to evaluate and compare the performance of different delivery policies. Furthermore, we derive a new state-dependent delivery policy that makes an optimal decision between partial or complete delivery for each retailer order that occurs. The policy is referred to as the MSD (Mixed State-Dependent) policy and it utilizes real-time information on when replenishments will arrive to the central warehouse. Given the MSD policy, we provide methods for exact cost evaluation and optimization of all reorder points in the system. We also provide exact methods for the pure partial delivery policy and the pure complete delivery policy, thereby providing alternative solution methods to problems that are solved in Andersson (1999), Axsäter (2000), and Forsberg (1997).

The analysis is based on utilizing the properties of a stochastic variable referred to as the nominal inventory position (denoted by $\Psi_0(\tau)$). Let L_0 be the central warehouse replenishment lead time, and let $D_0(0,\tau)$ denote the total amount of demand to the central warehouse in an arbitrary time interval $[0,\tau]$. Furthermore, let $IP_0(0)$ be the central warehouse inventory position at the beginning of the time interval. We then define the nominal inventory position at time τ ($0 \leq \tau \leq L_0$) as

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

The nominal inventory position is a stepwise decreasing variable that contains information about how much demand the central warehouse can satisfy before time L_0. Based on its properties we devise a scheme for separating the analysis into three mutually exclusive and collectively exhaustive events. By conditioning on these events we then obtain the probabilities needed to determine all performance measures (e.g., the distributions of all inventory levels). In the analysis we also obtain some analytical results regarding the different delivery policies that we consider. We show that the MSD policy has a performance guarantee compared to both the partial delivery policy and the complete delivery policy. In addition, we derive sufficient conditions for when it is always optimal to choose complete deliveries. This means that we can identify systems where the MSD policy will be identical to the complete delivery policy.

We also conduct a numerical study that consists of 32 problem scenarios. In this study, the expected total costs given the MSD policy are used as the benchmark. The results indicate that the relative cost increase of using the other simpler policies can be significant. Over the 32 scenarios, the average and maximum cost increases for the partial delivery policy were 5.8% and 26.6%, respectively. The corresponding values for the complete delivery policy were 5.9% and 17.9%. Our study also indicates that, given our new MSD policy, it is optimal to allocate more stock to the central warehouse than the previous literature on the subject suggests (see, e.g., Axsäter, 2003).

Thus our results show that it is important to choose the right delivery policy in the considered system. They also show that the value of using our new MSD policy can be significant.

3.4 Paper IV – A Distribution Inventory Model with Transshipments from a Support Warehouse

In paper IV we study a system with two supply options for the lower echelon. The system consists of a number of retailers, and a so-called support warehouse (Figure 12). The retailers face customer demand, while the purpose of the support warehouse is to provide additional demand coverage in case of stock-outs.

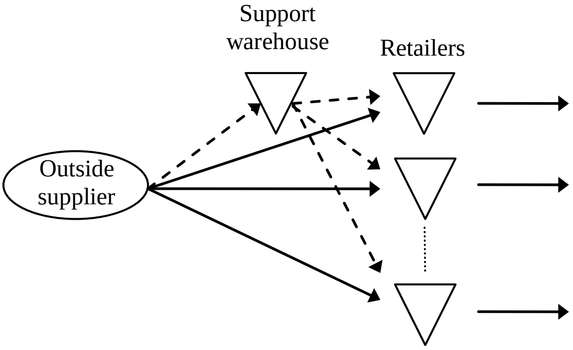

Figure 12. The system considered

More specifically, the retailers place replenishment orders to an outside supplier using (R,Q) policies. If demand occurs at a specific retailer and that retailer is out of stock, a transshipment (also known as an emergency shipment) for the amount needed to fulfill the demand is requested from the support warehouse. Customer demand is assumed to follow compound Poisson processes. However, our analytical model is based on approximating the lead time demands with normal distributions. The support warehouse, facing the demand for transshipments from the retailers, also uses an (R,Q) policy to place orders to the outside supplier.

We assume that the transshipment times are negligible, implying that demand is satisfied instantly when a request for transshipment is made, if the support warehouse has sufficient stock on hand. If, however, the support warehouse is out of stock, the request is backordered at this location and is then fulfilled upon the arrival of a replenishment order from the outside supplier. These backorders at the support warehouse are satisfied according to the First Come - First Served rule.

The cost structure consists of inventory holding costs per unit and time unit at all locations and penalty costs per unit transshipped from the support warehouse. Furthermore, the retailers must fulfill given service constraints. The service measure for each retailer (referred to as the combined fill rate) is the fraction of demand satisfied, either directly from stock on hand at that retailer, or from immediate transshipment from the support warehouse. Hence, target service levels can be achieved at a specific retailer by either keeping stock at the retailer, or keeping stock at the support warehouse and requesting transshipments, or a combination of both.

The motivation for analyzing this type of inventory system stems from research collaboration with Volvo Parts Corporation, a global after market service provider with headquarters in Sweden. Volvo Parts is responsible for the world-wide distribution of spare parts for vehicles and engines made by the Volvo group (including: Volvo Trucks, Mack, Renault Trucks, Volvo Buses, Volvo Construction Equipment, Volvo Penta and Volvo Aero). It follows that Volvo Parts has an extensive spare parts distribution system consisting of several central warehouses, positioned around the world. These central warehouses serve local markets where the support warehouse system that we consider is used. Given fixed order quantities, the objective with our model is to find the support warehouse and retailer reorder points that minimize system costs under the given service constraints, thus helping Volvo Parts to achieve a better coordination in their inventory systems. The model is designed to be computationally fast enough to handle large scale systems (on some markets there are over 100 retailers).

Exact analysis of the system considered is very complex due to the dependencies between the different inventory locations. Therefore, the evaluation of expected costs for given policy parameters incorporates some approximations. The cost evaluation is based on the observation that transshipments from the support warehouse can be viewed as demand being lost for the retailers, and transferred to the support warehouse. This enables us to decompose a complex multi-echelon problem into more manageable single-echelon lost sales problems. The mentioned approximations apply to the demand distributions used and the service level determination at the support warehouse. Our analysis also assumes that there can be at most one order outstanding with the outside supplier, or equivalently, $Q_j > R_j$ at each retailer j.

For finding the best reorder points we provide a heuristic that enumerates over given service levels at the support warehouse. As mentioned previously, the heuristic is computationally very efficient, making our model tractable for large real-life systems.

Because our analysis requires some approximations, we use simulation to validate that our model produces good results, that is, optimal or near-optimal values for the reorder points. In the simulations, customer demand is assumed to follow compound Poisson processes. Results from a numerical study show that our model is accurate as long as the assumption of at most one order outstanding is not violated, or the variance of customer demand is not too high compared to the mean. The former is of less practical importance, since it is rare that a retailer has more than one order outstanding in Volvo Parts' inventory systems. The latter is an expected consequence of using the normal distribution as an approximation for compound Poisson demand.

In a second numerical study we focus on the value of having a support warehouse. We compare the costs of the current setup with those resulting from a system where the retailers backorder all demand, instead of requesting transshipments from a support warehouse. Results show that utilizing a support warehouse structure can make sense, even if the costs of transshipments are high compared to the holding costs.

In a third study we compare the costs of using lateral transshipments between all retailers instead of using the support warehouse. Results indicate that applying lateral transshipments can be advantageous as long as the transshipment costs are the same in both systems. However, if transshipments from the support warehouse are cheaper; using a support warehouse can be preferable. Lower transshipment costs in the support warehouse system can be argued because of scale effects and higher efficiency in transshipment operations. Effectively, Volvo Parts has chosen the current support warehouse structure in order to have full control over transshipment operations and to avoid complicating incentive issues which typically arise when different independent firms share inventories.

Finally, we provide a numerical study based on data that was provided by Volvo Parts. We consider 50 representative articles involving 63 retailers on the Spanish market. For these articles, the solutions provided by our model are compared to the solutions currently used at Volvo Parts in a simulation study. Averaging over the 50 articles, our model reduces costs by 29%, while still achieving target service levels. Hence, the study indicates that there is a large potential for cost savings in applying our model to Volvo Parts’ inventory systems.

3.5 Paper V – Using Pipeline Information in a Multi-echelon Spare Parts Inventory System

The research in Paper V is also motivated by the collaboration with Volvo Parts Corporation. Once again we consider the support warehouse structure, but under different assumptions than in Paper IV. The objective is now to develop a policy that is more flexible in regards to the use of emergency shipments (transshipments). We focus on the low demand items in Volvo Parts’ product assortment, where there is an option to requesting emergency shipments from the central warehouse, in addition to using the support warehouse (Figure 13). This option was not considered in Paper IV, which focused solely on the dynamics between the support warehouse and the retailers (which are referred to as local warehouses in the current paper). Therefore, in the current paper the central warehouse affects the system through the emergency shipments and replenishment lead times that it provides.

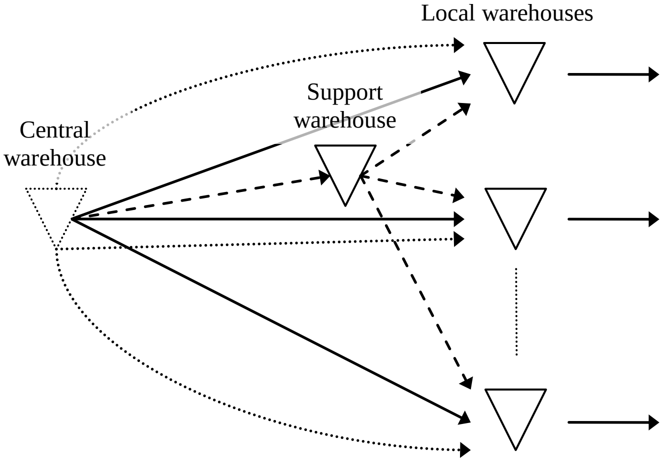

Figure 13. The system considered

It can, however, still be regarded as an external supplier because we maintain the assumption that it can always deliver on time (which is reasonable considering that the service levels at the central warehouse are currently very high in the real system).

Focusing on low demand items we assume that the order quantities are equal to one at all stock points, and that customer demand follows Poisson processes. When demand occurs at a local warehouse j, the warehouse applies a new policy, which uses real-time information on outstanding orders. The policy is referred to as an (S_j,T_j) policy. The decision variable S_j corresponds to the base-stock level at local warehouse j, while the decision variable T_j (referred to as the threshold time) determines if an emergency shipment should be requested in case of stock-outs. That is, if a stock-out occurs, items in the replenishment pipeline closer than T_j time units are reserved, and the demand is backordered. However, in the case of no unreserved items within reach of T_j time units, an emergency shipment is requested from the support warehouse. The support warehouse, in turn, applies the same policy, based on its own threshold time T_0: (i) if possible, satisfy the request from stock on hand and send an emergency shipment immediately, (ii) backorder the demand in anticipation of an incoming order and send an emergency shipment when the item arrives in stock, or (iii) deny the request for an emergency shipment. In case of (iii), the local warehouse requests an emergency shipment from the central warehouse instead.

In this system we consider customer waiting costs per unit and time unit instead of service constraints. As in Paper IV, we also have inventory holding costs per unit and time unit and emergency shipment costs per unit.

For the type of spare parts considered, Volvo Parts' general policy is to always ask for an emergency shipment when a stock-out occurs. However, as recognized by the company, this is not necessarily the best strategy in terms of cost efficiency. If there is an outstanding order arriving in the near future, it might be better to wait for this order, instead of requesting an emergency shipment. By optimizing the (S_j,T_j) policy parameters we can determine when it is reasonable to wait for the outstanding order. We can also evaluate the benefits of including real-time information on the location of outstanding orders (referred to as pipeline information) in this way. Furthermore, the optimal T_j-values also provide insights into suitable system structures for different articles. This is because the situation with all T_j-values set to zero corresponds to the current policy at Volvo Parts, where emergency shipments are always requested when stock-outs occur. Conversely, if all T_j values are equal to the replenishments lead times all unmet demand will be backordered, and the support warehouse and central warehouse will never be used for emergency shipments. The model can therefore be used strategically to determine which items should be stocked at the support warehouse, and for which items the central warehouse should supply emergency shipments. This is an important question for Volvo Parts.

In the multi-echelon literature on emergency shipments, it is often recognized that ignoring orders in the replenishment pipeline is suboptimal. Despite this fact, literature that incorporates pipeline information and includes the option of waiting for regular replenishment is scarce (see Paterson et al., 2011 for a general overview of transshipment models). It follows that, to the best of our knowledge, there are no previous multi-echelon models that consider the same policy as the one presented in this paper. However, there are some papers that deal with a related issue, namely, lateral transshipments with non-negligible lead times (Yang et al., 2012 and references therein).

Our analysis is based on first studying a single local warehouse in isolation. This single-echelon system is modeled as a queuing network, and by exploring the similarities to a dual supplier model (analyzed in Song and Zipkin, 2009), we provide an exact method for cost evaluation and optimization of the policy parameters S_j and T_j. We then utilize these results to derive a heuristic for setting base-stock levels and threshold times for the support warehouse and all local warehouses. The heuristic is evaluated in a simulation study, where it is shown to produce optimal or near-optimal solutions.

To investigate the value of incorporating pipeline information in the inventory control method, we present two numerical studies. The first study, encompassing 4800 problem scenarios, focuses on the penalty (i.e., the relative cost increase) of ignoring pipeline information at a single local warehouse. The main results obtained from this study are that the penalty appears to be increasing in the emergency shipment cost and holding cost, and that it is generally difficult to predict the properties of the optimal (S,T) policy for a given scenario. The second study considers the multi-echelon model and features 70 representative articles from Volvo Parts' inventory system in Spain. The study indicates that ignoring pipeline information can be quite costly, rendering an average penalty of 15% and a maximum penalty of 106%. Our study also indicates that using pipeline information in a simple way and always choosing the quickest replenishment option, can result in poor cost performance (penalties as high as 91% were recorded).

From the results obtained in this paper, we conclude that there is a large potential for increased cost efficiency by using pipeline information. Furthermore, the model is shown to be a viable decision support tool for determining how much to stock, where to allocate emergency supplies, and when to request an emergency shipment.

4. Contributions and future research

In this chapter we discuss the main contributions of each paper and provide an outlook on possible future research directions.

Paper I and Paper II compare the simple FCFS allocation rule to more advanced methods of allocation in a system with shipment consolidation. From a practical viewpoint, it is obviously easier to implement the FCFS policy than the MAs, MAd, HAs or HAd policy. It is also clear that allocation at the moment of delivery requires a very sophisticated information and distribution system, whereas allocation at the moment of shipment most likely could be applied to a broader range of systems. Thus, for advocators of the FCFS policy it is encouraging that this policy in many cases performs well compared to allocation at the moment of shipment. However, knowing that there are policies that perform better (such as MAd and HAd) might also serve as a motivation for developing more advanced information and distribution systems. From a theoretical viewpoint, our work holds importance because most of the exact analysis of continuous review multi-echelon distribution systems is based on the FCFS assumption. In many of these systems, allocation at the moment of delivery would not be feasible. Consequently, our work serves as a validation that the FCFS assumption often is reasonable.

We identify two main research directions that would be interesting to follow based on Paper I and Paper II. The first direction is to continue to generalize the policies considered. This could include more general demand processes (such as compound Poisson demand) and more general ordering policies (such as (R,Q) policies) at the retailers. The second direction more directly concerns the evaluation of the FCFS policy. In this work we compare FCFS allocation to feasible allocation policies. An alternative is to develop a lower bound for the expected total costs to offer insights on how FCFS might perform compared to the optimal policy. It would also be interesting to focus more closely on the determination of policy parameters. We apply the common approach of using the optimal FCFS parameters. Extending the limited study of simulation based parameter optimization in Paper I might give additional insights into the validity of this approach.

In Paper III we contribute to the stream of literature concerning exact analysis of two-level distribution systems. It is our hope that the technical contribution (i.e., the approach based on the nominal inventory position) will prove useful when analyzing other types of systems or problems. Furthermore, the work in Paper III highlights the importance of choosing the right delivery policy at the central warehouse. Our numerical study exemplifies how the choice between partial or complete deliveries can have a major affect on system costs. For example, assuming a partial delivery policy when optimizing reorder points may lead to suboptimal solutions where too little stock is kept at the central warehouse.

One of the main strengths of the new MSD policy is that it will never perform worse than any of the simpler policies considered. Hence, it is an example of how real-time information can be used to achieve cost benefits. At the same time, our tests suggest that if one applies the best of the two simpler policies, the loss of efficiency compared to the MSD might not be that high. This may suggest that real-time information is not needed to reap some of the benefits from a more careful selection of delivery policy. However, a larger study is needed before such claims can be made.

Regarding further numerical tests and future research, there are a number of interesting ways to experiment with the delivery policies that we consider. Clearly, it is possible to apply and evaluate any of our delivery policies, regardless of how the reorder points in the system are determined. Thus, one could, for instance, develop a fast approximate method for determining the reorder points and then evaluate the benefits of applying the MSD policy given these reorder points. Another obvious future research direction would be to generalize our method to compound Poisson demand.

Paper IV analyses Volvo Parts Corporation’s inventory systems with the objective to achieve coordination and efficient control. The main contribution is a heuristic method for determining reorder points which is capable of handling large scale real-life systems. The results obtained from data provided by Volvo Parts illustrate that the model can bring large cost savings. Furthermore, the type of inventory system studied is not unique for Volvo Parts. It is, for instance, also utilized by some of their competitors. Hence, the generality of this work is quite high. In regards to more theoretical contributions, previous papers on similar system structures with emergency shipments or lateral transshipments often focus on base-stock replenishment policies. It is also common to assume exponentially distributed lead times. Although our model is approximate, we provide more general and realistic assumptions in some aspects of the modeling, such as batch ordering policies and constant replenishment lead times. Moreover, our numerical experiments highlight some of the pros and cons with different types of system structures and when it might be reasonable to use a support warehouse structure.

Future research could include enhancements of some of the approximations used. A better approximation for handling the cases with multiple orders outstanding could, for instance, improve the robustness of the model. Another interesting challenge would be to make the model’s accuracy less sensitive to high customer demand variances, without compromising the computational tractability. Deriving a lower bound on the costs may also increase the validity of the approximations used. Moreover, from a structural perspective, including the inventory decisions at the central warehouse is a natural extension to Paper IV, although this is likely to require further approximations.

In Paper V we study the use of emergency shipments by developing a policy that utilizes real-time pipeline information. With examples from Volvo Parts we demonstrate the practical usefulness of the considered model for low demand items. However, we believe that the model could be of strategic value for high demand items as well. This is because the model can be used to identify suitable system structures and strategies for requesting emergency shipments. The main contribution to the current literature is that we show that the common policy of always requesting an emergency shipment when a stock-out occurs, can lead to poor cost performance. Furthermore, we highlight that using real-time information in simple ways, such as always choosing the quickest replenishment option, can also be far from optimal. The derivation of an exact method for evaluating the single-echelon system and using this as a building block for the multi-echelon system is, in our opinion, also a valuable technical contribution.

Future research directions based on Paper V could be to include the inventory decisions at the central warehouse, consider direct customer demand at the support warehouse and to include multiple support warehouses. These additions would make the system structure more general, but it would also make the analysis more complicated. It would also be interesting to see how the results on using real-time pipeline information carry over to batch ordering policies. This, however, would require a different method of analysis than the one used in Paper V.

In closing, the objective of the research presented in this thesis has been to develop mathematical models and policies for efficient control and increased understanding of stochastic multi-echelon inventory systems, with a focus on allocation decisions and the use of real-time information. In the five papers on which this thesis is based we have provided new methods of analysis, models and insights, relevant to both theory and practice. We have developed new policies that use real-time information such as: the MAs, MAd, HAs, HAd allocation policies, the MSD delivery policy, and the (S,T) ordering policy. Through several studies we have demonstrated that there is a large potential in using real-time inventory information, but we have also illustrated that the exact level of performance gains are largely context dependent. Thus, this thesis contributes to the existing literature and body of knowledge concerning allocation decisions, real-time information and efficient control of multi-echelon inventory systems.

References

Andersson, J. 1999. Coordinated multi-stage inventory systems with stochastic demand. Doctoral Thesis, Lund University.

Arslan, H., S.C. Graves, T.A. 2007. Roemer. A single-product inventory model for multiple demand classes. *Management Science* **53** 1486-1500.

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. 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.

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

Chen, F. 1999. On (R,nQ) policies in serial systems. In S. Tayur et al., Eds, *Quantitative models for supply chain management.* Kluwer Academic Publishers 73-109.

Chen, F. 2000. Optimal policies for multi-echelon inventory problems with batch ordering. *Operations Research* **48** 376-389.

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.

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

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

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

Hadley, G., T.M. Whitin. 1963. *Analysis of inventory systems*. Prentice-Hall Inc. Englewood Cliffs N.J.

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

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

Kleijn, M.J., R. Dekker. 1998. An overview of inventory systems with several demand classes. Econometric Institute Report 9838/A, Erasmus University, Rotterdam, The Netherlands.

Kranenburg, A. A., G. J. van Houtumn. 2009. A new partial pooling structure for spare parts networks. *European Journal of Operational Research* **199** 908-921.

Lee, H. L., S. Nahmias. 1993. Single-item, single location models, in S. C. Graves et al., Eds., *Handbooks in OR & MS* **4** 3-55.

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

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

Minner, S. 2003. Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics* **81-82** 265-279.

Moinzadeh, K., C.P. Schmidt. 1991. An (S-1,S) inventory system with emergency orders. *Operations Research* **39** 308-321.

Muharremoglu, A., Tsitsiklis, J.N. 2008. A single-unit decomposition approach to multiechelon inventory systems. *Operations Research* **56** 1089-1103.

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. 2011. Optimizing replenishment intervals for two-echelon distribution systems with stochastic demand. Working paper Duke University and Chinese University of Hong Kong.

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

Song, J. S., P. Zipkin. 2009. Inventories with multiple supply sources and networks of queues with overflow bypasses. *Management Science* **55** 362-372.

Teunter, R.H., W.K.K. Haneveld. 2008. Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering. *European Journal of Operational Research* **190** 156-178.

Veeraraghavan, S., Scheller-Wolf, A. 2008. Now or later: A simple policy for effective dual sourcing in capacitated systems. *Operation Research* **54** 850-864.

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.

Yang, G, R., R. Dekker, A.F. Gabor, S. Axsäter. 2012. Service parts inventory control with lateral transshipments and pipeline stock flexibility. Working paper Erasmus University Rotterdam and Lund University.

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.

Zheng, Y. S., A. Federgruen. 1991. Finding optimal (s,S) policies is about as simple as evaluating a single policy. *Operations Research* **39** 654-665.

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