LUND UNIVERSITY

Master of Science Thesis

MIOM05 Degree Project in Production Management

Multi-Echelon Inventory Optimization with Service Differentiation at Volvo Group

Faculty of Engineering

Division of Production Management

Authors: Oscar Frölund & Holger Hjortstam

Supervisor: Johan Marklund

Abstract

Title: Multi-Echelon Inventory Optimization with Service Differentiation at Volvo Group.

Authors: Oscar Frölund & Holger Hjortstam.

Supervisors: Johan Marklund, Lund University. Philip Mårtensson & Christian Beckers, Volvo Group.

Background: Within inventory control the distribution of spare parts poses multiple challenges. Being a critical area from a business perspective this is a field of great interest for Volvo Group. Currently the company utilizes a single-node optimization approach for the multi stage distribution network and they would like to investigate the potential value of using a holistic multi-echelon approach.

Purpose: This thesis project aims to investigate the impact of applying a recently developed service-differentiated multi-echelon omnichannel inventory control model in Volvos Service Market Logistics (SML) distribution network using discrete event simulation. More specifically, the study is based on a representative sample of items for the Group Trucks Operations (GTO) division from the Regional Distribution Center (RDC) in Johannesburg, South Africa.

Methodology: The methodology used in this master thesis consists of two main parts. Firstly, an exploratory research approach according to Höst et al. (2006) is conducted. This includes holding open interviews with selected Volvo employees, scrutinizing documentation and a literature review. Following this the second part of the methodology includes applying the four initial steps of an operational research framework presented by Hiller and Lieberman (2010). These steps include defining the problem, formulating and building an analytical model and testing the model through a numerical study.

Conclusions: The numerical study shows an average decrease of the expected inventory in the system by approximately 25 percent for the investigated items. This could be done while still on average maintaining the same fill rate as the currently used single-echelon model for a majority of the investigated items. In the cases where the fill rate is not met the negative deviation is small and lower than one percentage point on average in all cases.

Keywords: Service differentiation, multi echelon, omni channel, Inventory control, Supply Chain Management.

Preface

This master's thesis, conducted in collaboration with Service Market Logistics at Volvo Group, signifies the culmination of the authors' journey through the Supply Chain Management master program at Lund University's Faculty of Engineering LTH.

Firstly, we extend our sincere gratitude to our supervisor at Lund University, Prof. Johan Marklund, whose unwavering support and exceptional guidance have been invaluable throughout this thesis. We are also deeply appreciative of our supervisors at Volvo, Philip Mårtensson and Christian Beckers, for their warm welcome and continuous support from the outset. Furthermore we would also like to thank the staff at Service Market Logistics and Advanced Analytics at Volvo Group for their valuable insights when interviews were conducted at the company.

Additionally, we would like to express our thanks to all the personnel at the Division of Production Management at the Faculty of Engineering LTH, Lund University, where we dedicated the majority of our time to conducting this master's thesis. Their assistance and expertise have played a crucial role in the successful completion of this academic endeavor.

Oscar Frölund & Holger Hjortstam, Lund 2024

Table of Content

List of Figures

List of Tables

Abbreviations

Abbreviation:	Definition:
CDC	Central Distribution Center
DIM	Dealer Inventory Management
DSP	Dealer Stock Control Package
EM	Extended Model
EOQ	Economic Order Quantity
FCFS	First Come First Served
GTO	Group Trucks Operations
IL	Inventory Level
IP	Inventory Position
KPI	Key Performance Indicator
LPA	Logistics Partner Agreement
LT	Lead Time
OWMR	One-Warehouse-Multiple-Retailer
RDC	Regional Distribution Center
ROP	Reorder Point
SCM	Supply Chain Management
SDC	Support Distribution Center
SE	Single Echelon
SKU	Stock Keeping Unit

SML	Service Market Logistics
SS	Safety Stock
TSL	Target Service Level
VMI	Vendor Managed Inventory
VOR	Vehicle-Off-Road
VR	Virtual Retailer

Chapter 1. Introduction

In this section a background for the master thesis is provided. This includes an introduction to the field of Supply Chain Management and inventory control as well as an introduction to Volvo Group and the current distribution network and inventory control processes.

1.1 Background

For all organizations handling the flow of products or services within the organization referred to as Supply Chain Management (SCM) in an efficient way is an important issue. In a landscape where many organizations are truly global, acting in different countries, this issue becomes even more complex. The SCM concept encompasses all processes involved in transforming raw materials into finished goods or services provided to a customer (Axsäter, 2006). With a substantial impact on the financial output of an organization, SCM is an area of great interest for top management and if handled correctly this area could lead to a competitive advantage.

Inventory control systems are crucial tools in the field of supply chain management. They ensure that sufficient inventory is available in different parts of the organization to fulfill customer demands. These systems provide decision rules on how to manage inventories across the distribution network while minimizing the associated costs. The trade-off between the holding cost and customer service should be carefully revised to ensure that it is aligned with the aim of the organization.

Technological advancements have significantly impacted the field by enabling access to more information. This, in turn, has enhanced forecasting accuracy and provided comprehensive insights into supply chain operations. As a result of these advancements, system optimization has evolved. What was once primarily reliant on qualitative expertise and individual node optimization has now become much more intricate. Advanced systems can optimize the entire supply chain, offering a holistic approach.

The spare part industry, characterized by sporadic and low-demand patterns and a diverse array of articles, poses numerous challenges for inventory control systems. Often constituting a substantial portion of total sales over a product's lifecycle or as part of a contract, the distribution of spare parts demands close attention and strategic planning.

This thesis will be conducted for Volvo Group focusing on the inventory control of their spare parts. With a truly global organization active in nearly 200 markets the inventory control system for spare parts consists of many nodes in a complex structure (Volo Group, 2024a). The existing optimization of the system depends on single node optimization, focusing on individual warehouses. In this thesis the effects of implementing a holistic optimization of the inventory control system with service differentiation is to be investigated.

1.2 Volvo Group

Established in 1927 in Gothenburg, Sweden, the Volvo Group has become a prominent player in the global transportation industry. With a history spanning over 90 years, the company is recognized for its contributions to various sectors, including trucks, buses, construction equipment, and marine and industrial engines. Volvo Group is dedicated to providing reliable and sustainable transport solutions with a strong focus on safety (Volvo Group, 2024a). The group is divided into 13 divisions which is presented in Figure 1.

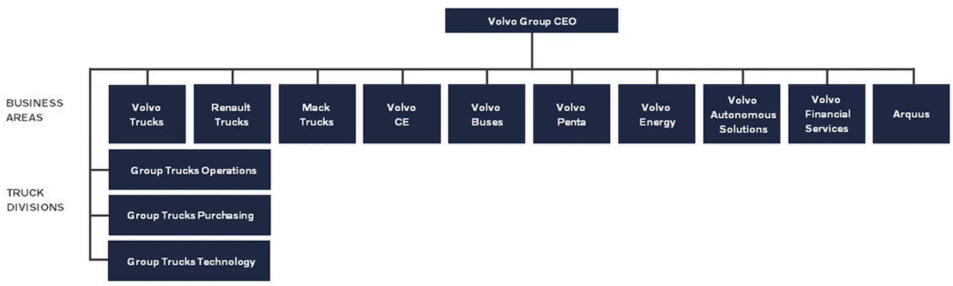

Figure 1: Volvo Group's corporate structure. (Volvo, 2024b)

This thesis will focus on the distribution of spare parts within the Group Trucks Operations (GTO) division of Volvo Group. Their inventory distribution system is managed by an administrative function called Service Market Logistics (SML).

1.3 The current distribution network and inventory control process

With a large organization to support, SML accounts for about 700 000 different spare parts[1]. The current distribution networks consist of several Central Distribution Centers (CDC), Regional Distribution Centers (RDC), Support Distribution Centers (SDC) showcased in Figure 2. The distribution network handles three main types of orders: stock orders, day orders, and vehicle-off-road (VOR) orders. These different types of orders follow a hierarchy of urgency. Stock orders refer to planned orders, day orders refer to emergency orders when stock is urgently needed and are sent the day after. Lastly, VOR refers to orders with severe urgency. These orders are sent directly to the customer from the closest stocking point instead of being treated as a day or stock order.

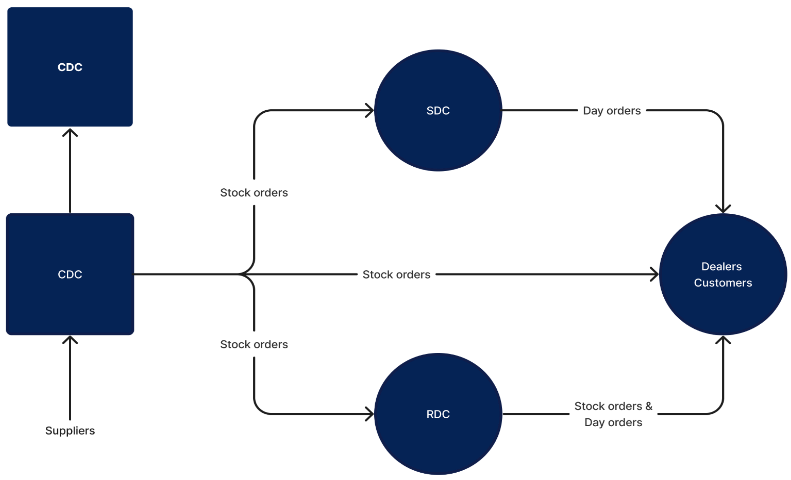

Figure 2: SMLs distribution network.

[1] Interviewee 1: Supply Chain Data Modelling Expert, Volvo GTO, SML, Advanced Analytics. Conducted February the 5th 2024.

In the current structure the CDCs supply the RDCs, SDCs and stock orders to the dealers which then serves the end customer. The distribution set-up however is different depending on the continent. For example in Europe where the dealer density is relatively high and homogeneous, the set up includes CDCs for stock orders and SDCs for day orders[2]. In other continents the set-up is normally less complex, consisting of a set of different RDCs where dealers order from a designated RDC. Volvo controls the inventory of several dealers through a Vendor Managed Inventory (VMI) agreement called Logistics Partner Agreement (LPA).It is noteworthy that some dealers with an LPA agreement are owned by Volvo, while others are not. For dealers with an LPA agreement that are not owned by Volvo, negotiations are required to establish inventory policies. Additionally, there are dealers not owned by Volvo and without an LPA agreement, whose inventory levels Volvo cannot dictate. These latter two categories of dealers will in this thesis be referred to as non-controllable or independent retailers, indicating that Volvo can not directly control their inventory decisions.

In the current inventory control system, Volvo uses a single-echelon modeling approach to calculate and optimize the inventory control decision. To be able to handle the great number of SKUs, products are segmented based on customer criticality, frequency, price, and life-cycle. The standard inventory policies used for replenishment decisions in the network are (R, Q) - policies. Once the inventory position (stock on hand and outstanding orders) reaches the reorder point (R), Q units are ordered. Q is determined using the Economic Order Quantity (EOQ)-model taking minimum and maximum order quantities and order handling costs into consideration. The reorder point is set as the lowest integer fulfilling the target service level. In chapter 2 Volvo's current inventory control system is explained in more detail.

[2] Interviewee 1: Supply Chain Data Modelling Expert, Volvo GTO, SML, Advanced Analytics. Conducted February the 5th 2024.

1.4 Problem formulation

The previously described distribution network and inventory control system developed throughout Volvo's history reflect the goal of being a reliable partner throughout the lifetime of the sold products. However, the development of distribution networks and control systems is an ongoing process to accurately represent the complex reality of distribution systems. An improvement in the inventory management system could potentially reduce the inventory levels while service levels are kept stable or even improved.

Considering the current distribution and inventory control system at Volvo, one promising area to analyze is whether a multi-echelon inventory management system with service differentiation would be suitable. Thus, this is the area where this master thesis will be conducted. One of the limitations of established multi-chelon models is when the distribution network also includes independent and non-controllable retailers. Traditionally, multi-echelon optimization models tend to push stock down-stream in order to achieve cost reduction and Target Service Levels (TSL). For the independent dealers, this is not a viable option. The inability to exert direct control over the stock levels requires a reevaluation of the conventional wisdom governing multi-echelon models, demanding a flexible framework to navigate the intricacies of service differentiation.

The lack of multi-echelon models accounting for service differentiation across multiple distribution channels introduces a multifaceted challenge in the formulation of effective inventory planning strategies. The problem formulation, therefore, revolves around devising approaches that not only acknowledge but actively accommodate for the variability of demand, meet differentiated service levels, allocation of stock, and minimize cost throughout the supply chain.

1.4.1 Purpose of the thesis

This thesis project aims to investigate the impact of applying a recently developed service-differentiated multi-echelon omnichannel inventory control model in Volvos Service Market Logistics (SML) distribution network using discrete event simulation. More specifically, the study is based on a representative sample of items for the Group Trucks Operations (GTO) division from the Regional Distribution Center (RDC) in Johannesburg, South Africa.

1.5 Delimitations

Due to being a large organization and the limited time of 20 weeks in which the master thesis should be conducted some delimitations have to be made. The thesis will focus on a smaller set of articles distributed from the Johannesburg RDC to dealers. The distribution network in this market consists of a number of different dealers replenishing from the RDC in Johannesburg, which in turn replenishes from the CDC in Gent, Belgium[3]. The scope is limited to the optimization of reorder points at the RDC and dealers within the region and not the distribution network as whole.

1.6 Structure of the report

In this section, the structure and disposition of this master thesis is described, providing a roadmap for the reader to navigate through the various chapters and sections:

1. Introduction: In this section, a background to the topic of supply chain management and inventory control system is provided together with a brief introduction of the case company. Additionally, the problem formulation, purpose of the thesis, and delimitations are stated.

2. Description of Volvo's Inventory Control system: Following the introduction, a more in-depth description of the current inventory control system at SML is provided to the reader. Here, the currently used segmentation processes and total cost model are described.

3. Methodology: This section presents the research framework used for structuring the work project. The methodology consists of two parts. Firstly,

[3] Interviewee 2: Dealer Inventory Management Analyst, Volvo GTO, SML, DIM. Conducted February the 21th 2024.

an explorative research approach was used. followed by a problem solving approach following the framework for Operations Research studies.

4. Theory: In this section, relevant theoretical concepts are discussed to provide the reader with a solid understanding of the inventory control system's parameters. Moreover, the section ends with a literature review examining the potential models satisfying the examined case company's requirements.

5. Inventory control modeling for spare parts at Volvo: Within this section, the chosen multi-echelon omnichannel inventory control model is described, and the considered heuristics are explained in more detail.

6. Numerical study: This section describes the numerical study conducted as part of this master thesis, including the data collection, analytical model, and the discrete event simulation model.

7. Result and analysis: In this section, the results and analysis derived from the numerical study are presented.

8. Conclusion: This section of the master thesis presents the conclusions of the completed work together with suggestions for future research topics within the field.

Chapter 2. Volvos inventory control system

Volvo Group employs a variety of digital systems to manage its supply chain across different divisions and markets. For the purposes of this master's thesis, the focus is on examining the Dealer Stock Control Package (DSP) system, used in the considered market. It is noteworthy that Volvo is in the process of developing a unified system called PlanIT, intended to merge all data systems. However, at present, PlanIT is only present at the RDC in Johannesburg for the investigated distribution network.

The inventory optimization process is an approach aimed at minimizing the expected costs within Volvo's spare parts distribution system. This process begins with a segmentation of the Stock Keeping Units (SKUs).

2.1 Segmentation

The department within SML that works with the segmentation is Dealer Inventory Management (DIM). The goal with the segmentation is to assign a dedicated stock and pick table for each SKU at the different dealers which in turn sets an reorder point and order quantity. In essence the process can be described by Figure 3 below.

In the first stage of the process, the investigated part and dealer are segmented separately. Here the dealer part file and dealer file contains data for the investigated part and dealer respectively. In the second stage, based on the outcome from step one, a stock and pick level version table is then assigned. Here the dealer segment prescribes if a high, normal or low stock/pick table version should be assigned. The SKU segment determines the appropriate categorization of the investigated item within the current version. In the final step, a stock/pick table is assigned which in turn sets the reorder point and EOQ. These steps will be described more thoroughly in the following sections.

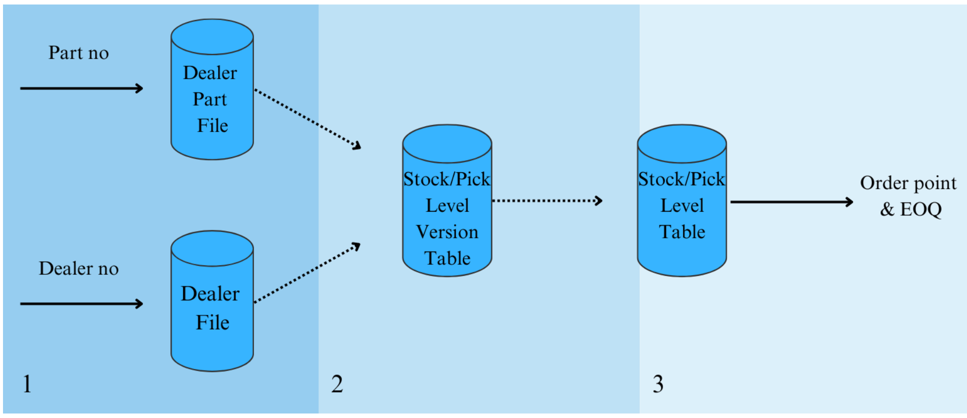

Figure 3: Segmentation within SML.

2.1.1 SKU and dealer segmentation

In the segmentation process, the SKUs are divided into different segments based on customer criticality, price, frequency and life cycle stage[4]. Additionally, heavy items, weighing over 30kg, are segmented into a separate group. Depending on where within the distribution network the segmentation is conducted the number of segments varies. For GTO the number of segments is normally 12 at the dealers which increases to 25 and 100 at the RDC and CDC respectively. This increase can be explained by the higher number of articles upstream in the distribution network. For the scope of this master thesis the focus is on the dealer segmentation. In Figure 4 below the segmentation for dealers in GTO is exemplified.

[4] Interviewee 1: Supply Chain Data Modelling Expert, Volvo GTO, SML, Advanced Analytics. Conducted February the 5th 2024.

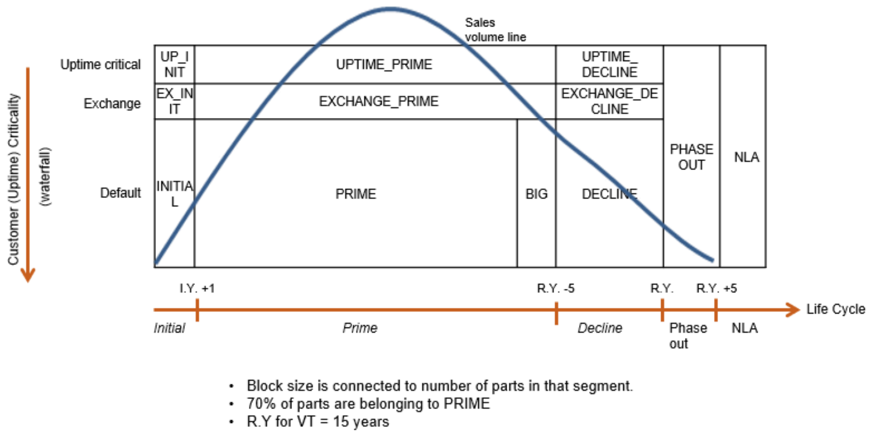

Figure 4: The 12 segment based on customer criticality and life cycle stage. (Volvo 2024c)

As stated in Figure 4, 70 percent of the SKUs belong to the PRIME segment. The life cycle axis uses two abbreviations; I.Y and R.Y which represents the introduction year and responsibility which represents the year when the SKU is expected to be phased out of the system. SKUs are segmented into the different stages according to Table 1 below.

For SKUs belonging to decline or phase out a second condition is investigating how many picks per year the SKU generates at the dealer. Here R12 denotes the rolling twelve pick which represents the number of picks during the last twelve months and C.Y which is an abbreviation for current year.

Table 1: Conditions for the different life cycle phases

Life cycle phase	**Condition 1**	**Condition 2**
Intitial	I.Y current or last year	
Prime	SKUs not belonging to another phase	
Decline	past R.Y <= C.Y +5	Hits R12 <= 15
Phase out	past R.Y <= C.Y -1	Hits R12 <= 6
NLA	past R.Y <= C.Y -6	Hits R12 <= 6

The dealers are segmented based on the size of the dealer and the business cycle. Here the business cycle indicates if the sales trend is upwards, stable or downwards and divides the dealers into high, normal or low respectively. The size of the dealer is based on the total number of rolling twelve picks.

2.1.2 Assigning a stock and pick table

Knowing the dealer segment it is possible to assign a stock and pick version table, see Figure 5. Here the version is indicated by the vertical text and divided into high, normal or low. The SKU segment then decides which table to be assigned to within this version.

For example a SKU that belongs to EX_INIT segment in Figure 4 is controlled with a high table if the dealer is assigned a high stock/pick version. However in the case of the dealer being assigned a medium or low stock/pick version the same SKU is assigned a normal table instead. Other SKU segments, for example UPTIME_PRIM or NLA, are controlled the same way regardless of the dealer type.

Version	Info Code	Table	Version	Info Code	Table	Version	Info Code	Table
High Stock/Pick Version	VO_BIG	Normal Table	Normal Stock/Pick Version	VO_BIG	Low Table	Low Stock/Pick Version	VO_BIG	Low Table
	VO_DEC	Low Table		VO_DEC	Low Table		VO_DEC	Low Table
	VO_EX_DEC	Low Table		VO_EX_DEC	Low Table		VO_EX_DEC	Low Table
	VO_EX_INIT	High Table		VO_EX_INIT	Normal Table		VO_EX_INIT	Normal Table
	VO_EX_PRIM	High Table		VO_EX_PRIM	Normal Table		VO_EX_PRIM	Low Table
	VO_INIT	High Table		VO_INIT	Normal Table		VO_INIT	Normal Table
	VO_NLA	Low Table		VO_NLA	Low Table		VO_NLA	Low Table
	VO_PHOUT	Low Table		VO_PHOUT	Low Table		VO_PHOUT	Low Table
	VO_PRIME	High Table		VO_PRIME	Normal Table		VO_PRIME	Low Table
	VO_UP_DEC	Normal Table		VO_UP_DEC	Normal Table		VO_UP_DEC	Normal Table
	VO_UP_INIT	High Table		VO_UP_INIT	High Table		VO_UP_INIT	High Table
	VO_UP_PRIM	High Table		VO_UP_PRIM	High Table		VO_UP_PRIM	High Table
		Normal Table			Normal Table			Low Table

Figure 5: Stock and pick table depending on segment and version. (Volvo 2024c)

The allocated stock and pick table then determines the order point and safety stock levels for the SKU at the specified dealer. Here it is price and forecast which decides on where in the table the parameters should be taken.

2.2 Total cost model

Knowing the associated stock and pick table it is now possible to optimize the parameters for the distribution network. At Volvo this is done with the total cost model, see Figure 6 below.

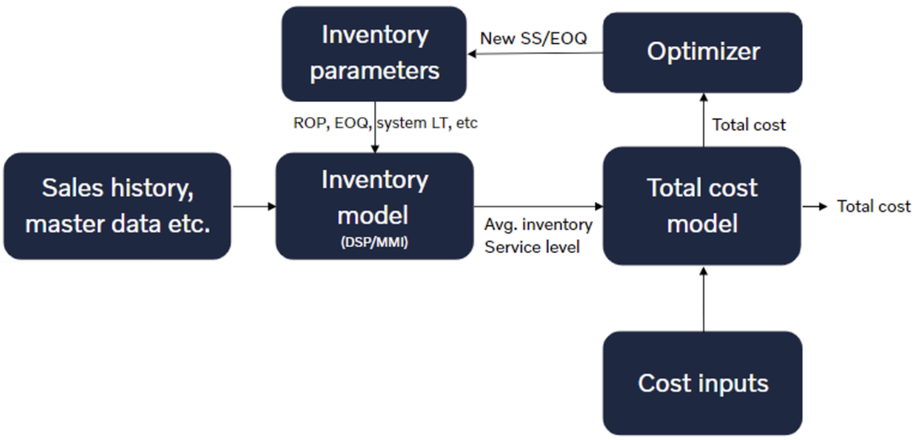

Figure 6: Overview of setting inventory parameters at SML. (Volvo 2024c)

In the total cost model the goal is to minimize the total cost under the given constraints. Starting at the inventory model this is where the actual distribution network is mimicked. Here input data comes from sales history, master data etc but also from the model itself including for example Reorder Point (ROP), EOQ and system Lead Time (LT). As previously mentioned there are different systems throughout the Volvo organization, we are focusing on DSP which is one of the systems used at Group Trucks Operations (GTO).

Within the DSP model, the distribution of demand for each segmented category of products is determined. This step is particularly crucial in determining the appropriate inventory management strategies for different product categories. In instances where the number of annual picks falls below a predefined threshold, the demand pattern is assumed to follow a pure Poisson process. In simpler terms, customers are considered to arrive at a certain rate (lambda) and are restricted to ordering only one product per transaction. Conversely, for orders surpassing this threshold, a normal distribution model is employed to predict the demand distribution. Notably, this modeling approach does not account for the specific order quantities but focuses on forecasting the overall customer arrival.

Subsequently, the DSP model utilizes the acquired parameters to compute key inventory metrics, including inventory on hand and service levels. These calculations are heavily influenced by the current settings of reorder points, EOQ, and safety stock levels defined within the DSP system. By leveraging these parameters, Volvo aims to strike a balance between maintaining optimal inventory levels to meet customer demand while minimizing expected total costs.

The output of the inventory model is then run through a total cost model. In this step cost inputs need to be provided which breaks down where the cost comes from. Here the cost is not only including direct holding and shortage costs of SKUs but also for example the cost associated with badwill or lost sales due to stockout. Volvo segment the cost inputs in different levels including for example country, segment, weight and brand.

To achieve the minimal total cost an optimizer can be run where the parameters are shifted, resulting in a situation with the lowest possible cost. Planners may opt to manually adjust critical parameters such as EOQ, safety stock levels, and reorder points to evaluate the potential impact on the overall cost of inventory management. In response to these adjustments, the DSP model recalculates service levels and total inventory, providing valuable feedback for decision-making purposes.

The total cost model can also be used to provide for base-line cases. This can be valuable when evaluating the impact of a change project. In this master thesis the base-line for evaluation will be a newly made optimization with the total cost model for the South African market.

Chapter 3. Methodology

This section outlines the methodology adopted in this thesis and consists of two studies. In section 3.1 an exploratory research approach study is conducted to understand the case company and the requirements on the inventory control system investigated in this thesis.

Section 3.2 presents a problem solving framework from operations research. This framework is used for evaluating the potential of using a multi-echelon/omni-channel inventory control model at Volvo. In this section, the foundational framework consisting of a six step approach, is firstly introduced as a whole. The initial four steps, chosen as methodology for this thesis, is then described in more detail and how this is taken into practice elaborated further.

3.1 Exploratory research approach

This segment of the research aligns with the framework of an exploratory research approach, as exemplified in (Höst et al., 2006). In the context of an exploratory study, they advocate for the utilization of a case study approach. The subject of interest is on the inventory control system and distribution network for a selected market within SML. Consistent with their recommendations, our study employed three techniques for data gathering:

The first technique consists of conducting open interviews with selected Volvo employees**.** The interviews will adopt a hybrid approach, encompassing elements of both structured and open interviews. To ensure a comprehensive exploration of the subject matter, a set of both open-ended and specific questions has been prepared, see Appendix A. These questions are designed to capture essential information while providing room for in-depth responses and facilitating follow-up questions. Within this framework, certain follow-up questions have been pre-established. Additionally, the interview process allows for follow-up questions as the conversation unfolds. This structure ensures a flexible and thorough examination of the subject, catering to the expertise and insights offered by the interviewee.

The second technique includes scrutinizing documentation, including technical documents, IT systems and process guides: Through a thorough research of technical documents, valuable insights into the system's architecture and design are obtained. Process guides provide an in-depth exploration of operational procedures, enabling the identification and definition of processes relevant to distribution and inventory control. The research of the IT system focuses on acquiring essential data necessary for the formulation of a mathematical model, encompassing factors like transportation times, service requirements, and demand.

The third technique is analyzing the process steps: Conducting an analysis of the information obtained in the two previous steps allows for a comprehensive understanding of the distribution and inventory control system at Volvo. The required information to establish a mathematical model is presented in Table 2, and the output establishes the criteria that the models investigated in the literature review must meet to accurately represent the system at Volvo.

Table 2: Important data for distribution system modeling.

Elements	**Description**
Network specifications	● Number of echelons and nodes in the network ● The flow of goods between echelons and nodes ● relationship between nodes (divergent or convergent distribution)
Market specifications	● Which dealers and inventory policies Volvo can control ● Reaction to missed order (back-ordering, lost sales, partial deliveries) ● Service level target for each product and dealer ● Capacity constraints
Detailed specifications	● Transportation times between nodes ● Inventory policies ● Demand distribution for each product ● Cost structure (holding costs and backorder/lost sales costs per product)

3.1.1 Literature review

The literature review aims to establish a comprehensive understanding of the concepts and tailored applications relevant to the intended scope of the study. For this study, Höst et al. recommends: search wide, select relevant articles, and search deep. (Höst et al., 2006)

Search Wide: To gather a collection of relevant articles, various strategies were employed. These methods encompassed seeking recommendations from our supervisor, particularly for articles focused on multi-echelon models with service differentiation. Additionally, we conducted a thorough examination of reference lists from previous related theses, utilized the university database Lubsearch, and scrutinized reference lists from a range of relevant articles.

To initially search wide, the keywords used were *service differentiation AND inventory Control*. This search resulted in 249 publications on LubSearch. The timespan was set from 1960 to 2024, after dialog with the master thesis supervisor prof. Johan Marklund.

Select relevant articles: The choice of papers was determined by their relevance to the master thesis and the quality of the content. The assessment of relevance centered on whether the researcher proposed a mathematical model to accommodate service differentiation and provided a numerical study demonstrating its effect. Subsequently, the articles were evaluated based on quality using the criteria suggested by Höst et al. (2006), which include considerations such as the number of citations, if the paper has been peer-reviewed, and the scientific soundness of the methodology.

Assessing the quality of the initial search and limiting the search to academic journals and peer-reviewed publications, the initial search was reduced to 201 articles.

Search Deep: Drawing insights from the examined papers, several key areas emerge as significant for further investigation, serving as supplementary knowledge and further alignment with the purpose of the thesis. This resulted in an expanded search with keywords: *Inventory rationing AND inventory control, Service differentiation AND multi-echelon, Stock allocation AND multi-echelon*.

3.2 An operations research framework

The framework employed in this study is rooted in the general six step framework for operations research studies in Hiller and Lieberman (2010). The general framework is presented below. For our thesis, a slightly modified version of this approach is applied to better suit the purpose and scope of the project.

Step 1: Define the problem of interest and gather relevant data: Establishes the foundation for the project, encompassing essential actions such as formulating a precise problem definition, identifying relevant objectives and constraints, gathering relevant information, and establishing a timeline for the project's execution.

Step 2: Formulate a mathematical model: The formulation process involves translating the real-world problem into a system of equations and related mathematical expressions. This includes setting relevant parameters, decision variables, constraints, and objective functions.

Step 3: Learn how to derive solutions from the model: After formulating a mathematical model, the subsequent phase involves developing a procedure to find optimal solutions. It's crucial to recognize that these optimal solutions are specific to the model and may not guarantee the best possible outcome for the real-world problem, given the idealized nature of the model. Post Optimality analysis, also known as what-if analysis, becomes paramount in assessing the implications of different assumptions about future conditions on the optimal solution. Sensitivity analysis helps identify critical parameters in the model, whose values significantly impact the solutions.

Step 4: Test the model: This process involves testing and improving the model to increase its validity. Additional insights into the validity of the model can be obtained by varying the values of the parameters and/or decision variables and checking whether the output of the model behaves in a plausible manner.

Step 5: Prepare to apply the model: This step entails the comprehensive development of a computer-based system and the necessary business processes essential for the effective implementation of the newly devised model

Step 6: Implementation: The final phase is to execute this system in accordance with the directives from management. This process encompasses various steps, such as educating operational management and fostering shared responsibility for formulating the necessary procedures to set the system into operation.

For the scope of this master thesis, the focus is on step 1 - 4.

3.2.1 Step 1 - Define the problem of interest and gather relevant data

The exploratory study lays the groundwork by gathering relevant information about the case company's distribution and inventory system and pinpointing the challenges that the model under investigation in this master thesis should address. Following this, the literature review investigates models capable of addressing these challenges. Consequently, a model mirroring the real-world distribution system conditions is chosen for testing. The primary objective is to evaluate this model's performance to attain target service levels while minimizing stock

3.2.2 Step 2 and 3 - Formulate a mathematical model and develop an analytical model

After establishing a theoretical foundation and confirming its applicability, the next phase is constructing a mathematical model The mathematical modeling process includes identifying system parameters, decision variables, objective functions for optimization, and formulating mathematical constraints (Hiller and Lieberman, 2010). Once the mathematical model is established, the next step is to program the model in Python according to Volvo's specifications. The implementation of the model proceeds incrementally, with continuous testing to ensure verification. To validate the analytical model, a comparative analysis is conducted against a corresponding model built in Excel by Prof. Johan Marklund for a scaled-down system. This comparison ensures that the Python-based model produces identical results.

3.2.3 Step 4 - Test the model

To evaluate the analytical model and assess the effects, a discrete event simulation will be developed in the software ExtendSim 10 used by Volvo. The process of conducting a discrete event simulation study in ExtendSim adheres to the framework proposed by Laguna and Marklund (2018). This approach ensures a systematic and comprehensive examination of the inventory management system's performance, allowing for a more accurate understanding of its real-world implications.

3.2.3.1 Verification and validation

Model verification focuses on confirming that the model is implemented correctly and functions as intended. This involves ensuring that the model's algorithms and code are free from errors. Verification includes activities such as inspecting the model's code for logical errors or bugs, running tests on individual components to ensure each part functions correctly, confirming that the model produces consistent results when run under the same conditions, and ensuring that the model's outputs match expected results or results from previously validated models. (Laguna and Marklund 2018)

Model validation involves confirming that the model effectively achieves its intended purpose. This typically entails verifying that the model accurately predicts outcomes under the conditions for which it was designed. Validation techniques include comparing the model's predictions with actual outcomes, conducting sensitivity analysis to examine the model's robustness, and applying statistical methods to evaluate the model's performance. (Laguna and Marklund 2018)

3.2.3.2 Analyze the output data and draw conclusions.

In this final step, a thorough analysis of the output data from the simulation is conducted to derive meaningful insights and draw informed conclusions.

By simulating a range of scenarios, this step offers a holistic understanding of how the analytical model would function under diverse and dynamic real-world conditions. The simulation serves as an important tool for validating the robustness of the developed inventory management framework.

Chapter 4. Theory

This section presents relevant theory, for understanding the optimization of Volvo's inventory system for Service Market Logistics (SML). Section 4.1 to 4.3 focus is put on exploring and evaluating the single-echelon inventory control modeling area. This includes discussing fundamental concepts as well as diving into important areas such as the demand distribution and the lead-time approximation. A complete list of notations used in the formulas presented in these sections can be found in Appendix B.

Sections 4.4 and 4.5 introduce multi-echelon inventory control models, and the optimization of both single and multi-echelon is discussed. Here an emphasis is put on how the models can handle service differentiation since this is one model requirement from Volvo when investigating potential multi echelon models.

4.1 Fundamental inventory control concepts

In this section fundamental inventory control concepts are discussed in order to understand the currently adopted single node optimization at Volvo. Furthermore this lays the theoretical foundation necessary for the following sections of this master thesis.

4.1.1 Inventory control systems

Inventory control systems are crucial tools within supply chain management, serving to guarantee the presence of adequate inventory to meet customer demands. These systems not only establish decision-making rules for managing inventory throughout the distribution network but also aim to pinpoint strategies that optimize value while minimizing associated costs. Essential for the success of inventory control systems is the utilization of modeling techniques, which serve to accurately replicate real-world scenarios and provide a solid foundation for decision-making.

The models fall into two main categories: deterministic and stochastic. Deterministic models assume certainty, with fixed and constant values for parameters like demand, lead times, and order quantities. In contrast, stochastic models explicitly consider uncertainty, treating parameters as random variables with associated probability distributions.

4.1.2 Ordering systems

One fundamental aspect of inventory control systems is the ordering system, which refers to systems that provide the decision maker with information when and how much to order. Depending on what ordering and review policies the ordering system adopts, the information will vary. However, before delving into these policies, it is crucial to comprehend the fundamental concepts of inventory position (IP) and inventory level (IL) defined in (1) and (2) (Axsäter, 2006).

$$Inventory\ position\ = stock\ on\ hand\ + outstanding\ orders - backorders \quad (1)$$

$$Inventory\ level\ = stock\ on\ hand\ - backorders \quad (2)$$

In these functions, *stock on hand* is referred to as the physical stock in the warehouse at the moment of review. *Outstanding orders* refers to orders that have been placed with suppliers but which have not yet been delivered. *Backorders* are the units demanded that have not yet been satisfied by the stock point.

4.1.2.1 Ordering policies

Ordering policies refers to the standardized instructions of when and in which quantity a item should be replenished. The ordering policy typically relies on a specified order quantity triggered when the inventory position drops below a predetermined reorder point. In this section two commonly employed ordering policies: the (R,Q)-policy and the (s,S)-policy is discussed.

Under the (R,Q)-policy, an order quantity of Q units is placed as soon as the IP reaches or falls below the reorder point R. It may be necessary to order several Q to increase the IP above R and therefore also sometimes denoted to as (R,nQ) policy instead (Axsäter, 2006). This modified policy can also be used in the case of batchorders where Q is a fixed batch of units, e.g. a pallet, and n is the number of batches ordered.

On the other hand, the (s,S) policy, also known as the order-up-to or base-stock policy, triggers an order when the inventory reaches the reorder point s, up to the stock level S, for continuous and unit demand. Unlike the (R,Q)-policy, the order quantity fluctuates and cannot be predetermined. A notable instance of this policy occurs when s equals S - 1, indicating that an order is placed whenever demand is encountered (Axsäter, 2006).

4.1.2.2 Review policies

Review policies refer to the standardized instructions of how the system is monitored. The inventory system may be monitored continuously or periodically. If the system is monitored continuously the decided order quantity will be ordered as soon as the IP reaches the reorder point. If an (R,Q) policy is used, this can be formulated into an equation according to (3) (Axsäter, 2006). In continuous review, the R,Q policy and the s,S policy is equivalent if $s = R$ and $S = R + Q$.

$$IP \geq R + 1 \qquad (3)$$

In the case of a periodic review policy the inventory system is reviewed and orders can be placed only at certain times, often with a fixed time interval T between. As a direct consequence of this, the IP can fall below the reorder point before an order is placed, and equation (3) no longer holds.

In order to determine cost efficient reorder points and order quantities the lead time L to replenish stock from suppliers and the time interval T has to be taken into consideration if a periodical review policy is used. A continuous review policy normally leads to reduced need of safety stock (SS) however the periodical policy makes pooling of orders easier which can reduce transport costs especially for items with high demand. (Axsäter, 2006)

4.1.3 Inventory position and inventory level in steady state

Steady state refers to a situation where the distribution of stochastic variables do not change with time. Considering a system at the time *t* the IL can be calculated for the system at the time $t + L$ according to (4) (Axsäter, 2006). Here, $D(t,\ t + L)$ is the stochastic demand in the interval t, t+L and L is representing the replenishment lead time and is assumed to be constant.

$$IL(t\ +\ L)\ =\ IP(t)\ -\ D(t, t\ +\ L) \qquad (4)$$

In steady state, the relation must hold for every t and this time index can be discarded. Hence, (4) can be formulated as IL = IP - D(L) instead.

When dealing with an inventory system subject to stochastic demand, both the inventory position and the inventory level become stochastic (Axsäter, 2006). To determine the steady-state distribution of the inventory level, it is necessary to consider the distributions of both the inventory position and the lead time demand.

A system using a (R,Q) ordering policy with a discrete stochastic demand has an IP that is uniformly distributed on the integers R+1, R+2 …, R+Q in the steady state according to (5) (Axsäter, 2006).

$$IP \in U[R\ +\ 1,\ R\ +\ Q] \qquad (5)$$

4.1.4 Stock-on-hand and backorders

The concept of *stock-on-hand* and *backorders* was briefly discussed in section 4.1.2 However since these concepts come with an associated cost and thereby play a vital role in optimizing the inventory decisions this section will discuss the concepts more in depth.

The cost associated with holding stock-on-hand is often referred to as *holding cost (h).* These costs only occur when the IL is greater than zero (Axsäter, 2006). On the other hand, cost associated with backorders referred to as *backorder costs* (b_1) only occurs if the IL is negative (Axsäter, 2006). One way of describing the backorder cost is as a penalty cost per unit and time unit which the warehouse has to consider if the demand from the customers is not fulfilled straight away.

Naturally these two costs are connected, and the cost optimal solution involves balancing the two. If a continuous review (R,Q) policy is considered, the optimal solution minimizing cost can be calculated according to (9). Here it is convenient to use the notation in (6), (7) and (8). In (9) $hE[IL]^+$ denotes the expected holding cost and $b_1E[IL]^-$ denotes the expected backorder costs.

$$(x)^+ = max(x, 0) \qquad (6)$$

$$(x)^- = max(-x, 0) \qquad (7)$$

$$x^+ - x^- = x \qquad (8)$$

$$hE[IL]^+ + b_1E[IL^-] = -b_1IL + (h + b_1)E[IL]^+$$
$$= hIL + (h + b_1)E[IL]^- \quad (9)$$

4.1.4.1 Minimizing cost of (R,Q) - policies

Minimizing the cost of the (R,Q)-policy looking at holding and backorder cost can be done according to (10). In this function the holding cost is linear to the expected stock on hand and the expected backorder cost is linear to expected backorders. To get the correct optimal solution it is of greatest importance that these costs actually mimic reality. For example the backorder cost should include the loss of goodwill and future sales it brings if a customer can not be served straight away.

$$min_{R,Q}C(R, Q) = min_{R,Q}(hE[IL]^+ + b_1E[IL^-]) \qquad (10)$$

4.2 Demand during lead time

A cornerstone in the modeling process is understanding the demand; therefore, it is important to find a suitable demand model that characterizes the demand over a given time period (the replenishment lead time). In practice, the demand during a certain time is nearly always a nonnegative integer, i.e., it is a discrete stochastic variable (Axsäter, 2006). If the demand is reasonably low, it makes sense to use a discrete demand model to describe the real-world demand. In the case of relatively high demand, it makes more sense to use a continuous demand model as an approximation (Axsäter, 2006).

Since demand frequently occurs in discrete quantities rather than continuously, opting for a discrete demand distribution is a logical choice. Nevertheless, for computational efficiency, one might consider using a continuous distribution such as the Normal distribution. This section elaborates on some common options for distributions in this context.

4.2.1 Discrete demand: Compound Poisson distribution

A compound Poisson distribution represents the likelihood of the cumulative sum of independent and identically distributed incoming orders. Assuming a system where customers arrive according to a Poisson process with a mean arrival rate of λ customer per time unit. This means that the number of customer arriving during a time unit t is Poisson distributed with mean λ. The probability of k customers will follow a Poisson distribution according to (11). Note here that the expected time between customer arrivals is $\frac{1}{\lambda}$.

The probability that a single customer demands j units is denoted f_j and the probability that k customers having a total demand of j is f_j^k. This can be calculated recursively knowing that $f_0^0 = 1$ and $f_j^1 = f_j$ and is showcased in (12).

$$P(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \ k = 0, 1, 2... \qquad (11)$$

$$f_j^k = \sum_{i=k-1}^{j-1} f_i^{k-1} \cdot f_{j-i}, \ k = 2, 3, 4... \qquad (12)$$

Using (11) and (12), the stochastic demand in the time interval t, $D(t)$, can be calculated according to (13). A specific scenario arises when employing the compound distribution where $f_{j=1}$ equals 1, indicating that each customer consistently purchases only one item. In this case, the demand during the lead time t follows a Poisson distribution, offering an appealing estimate primarily because of its computational efficiency (Axsäter, 2006).

$$P(D(t) = j) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \cdot f_j^k \qquad (13)$$

The average demand μ per time unit can be determined using the function (14) provided below. In this context, K represents the stochastic number of arriving customers, Z denotes the overall demand, and J represents the stochastic demand size within the time unit. The preceding assumption that the arrivals follow a Poisson distribution enables the calculation of both the variance and the mean arrival rate according to (15).

$$\mu = E(Z) = E_K\{E(Z|K)\} = E_K\{K \cdot E(J)\} = E(K) \cdot E(J) = \lambda \sum_{j=1}^{\infty} j \cdot f_j$$
(14)

$$\sigma^2 = \lambda \cdot E[J^2] \rightarrow \lambda = \frac{\sigma^2}{E[J^2]} = \frac{\sigma^2}{\sum_{j=1}^{\infty} j^2 \cdot f_j} \qquad (15)$$

From (14) and (15) it is notable that the variance-to-mean rates ≥ 1, with equality only in the case of pure Poisson demand (Axsäter, 2006). Consequently, it is not possible to model the demand as compound Poisson demand if $\sigma^2/\mu < 1$. This constraint for using compound Poisson distribution can be expressed and evaluated numerically according to (16).

$$\frac{\sigma^2}{\mu} = \frac{\lambda \cdot E[J^2]}{\lambda \cdot E[J]} = \frac{E[J^2]}{E[J]} \geq 1 \qquad (16)$$

4.2.2 Discrete demand: Logarithmic compounding distributions

In practical scenarios, the shape of the demand distribution is often unknown and the distribution data limited. Nevertheless, when dealing with discrete, stochastic, and independent demand, a commonly employed assumption is that the demand size adheres to a logarithmic compound distribution according to (17). This choice is primarily made for computational simplicity.

$$f_j = -\frac{\alpha^j}{ln(1-\alpha) \cdot j} \qquad (17)$$

The parameters α, λ is computed according to (18) and (19), here μ' is a estimation of the mean value of the demand during the lead time (L), and σ' the estimated standard deviation of the demand during the lead time σ' (Axsäter, 2006).

$$\alpha = 1 - \frac{\mu'}{\sigma'^2},\ 0 < \alpha < 1 \qquad (18)$$

$$\lambda = -\frac{\mu'}{L}\frac{(1-\alpha) \cdot ln(1-\alpha)}{\alpha} \qquad (19)$$

Theory shows that the demand during the lead time for a logarithmic compounding distribution follows a negative binomial distribution according to (20) (Axsäter, 2006). If μ' and σ' is given the parameter p is equal to α and therefore calculated according to (18) and r is computed according to (21).

$$P(D(t) = k) = \frac{r(r+1)...(r+k-1)}{k!}(1-p)^r p^k,\ k = 1, 2, 3... \qquad (20)$$

$$r = \mu' \frac{(1-p)}{p},\ r > 0 \qquad (21)$$

4.2.3 Continuous demand: Normal distribution

In the case of high demand, regardless of the real distribution, a normal distribution approximation for the lead time demand is often a suitable option according to the central limit theorem, presuming the lead-time is sufficiently long. However a drawback of this approximation is the probability of negative lead time demand when the standard deviation is relatively high compared to the mean. (Axsäter, 2006)

For the normal distribution representing demand during the lead time, two parameters are necessary: the mean value during the lead time μ' and the standard deviation during the lead time σ'. These parameters are calculated according to equations (22) and (23), where L represents the lead time.

$$\mu' = \mu \cdot L \qquad (22)$$

$$\sigma' = \sqrt{Variance \cdot L} = \sqrt{\sigma^2 \cdot L} \qquad (23)$$

The lead time demand can be calculated using the density function and distribution function according to (24) and (25) (Axsäter, 2006).

$$f(d)_{D(L)} = \frac{1}{\sigma'\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot(\frac{d-\mu'}{\sigma'})} \qquad (24)$$

$$F(d)_{D(L)} = \int_{-\infty}^{d} f(x)_{D(L)} dx \qquad (25)$$

4.2.4 Continuous demand: Gamma distribution

In case the standard deviation-to-mean ratio ,$\frac{\sigma'}{\mu'}$, is high there is a chance of having negative demand if normal distribution approximation is made. An alternative to the normal distribution approximation is therefore to use a gamma distribution approximation which does not allow negative demand. This distribution has a density according to (26) where $\Gamma(r)$ denotes the gamma function and is computed according to (27) see, for example (Axsäter, 2006).

$$g(x) \ = \ \frac{\lambda(\lambda x)^{r-1}e^{-\lambda x}}{\Gamma(r)}, \ x \geq 0, \ \lambda, r \ > 0 \qquad (26)$$

$$\Gamma(r) \ = \ \int_0^\infty x^{r-1} e^{-x} dx \qquad (27)$$

The gamma distribution has a mean of $\frac{r}{\lambda}$ and a variance of $\frac{r}{\lambda^2}$. If μ' and σ' is given, r and λ can be computed according to (28) and (29).

$$r \ = \ (\frac{\mu'}{\sigma'})^2 \qquad (28)$$

$$\lambda \ = \ \frac{\mu'}{\sigma'^{\ 2}} \qquad (29)$$

4.2.5 Stochastic lead times

The lead time is defined as *the time between placing an order and receiving it.* To achieve computational simplicity it is common to use a constant lead time. However in reality the lead time is more likely to be stochastic. The processes included in delivering an order depend on several factors, for example, the inventory level at the upstream warehouse satisfying the order and the time transporting the specific order. Axsäter (2006) describes two types of stochastic lead times: *independent lead times* and *sequential deliveries independent of lead time demand.*

In the case of sequential deliveries independent of lead time demand, orders cannot cross each other in time. This type of stochastic lead time is the most commonly used in practice. In a model using this definition of lead times, a first-come, first-served policy is employed, implying that the customer ordering first is also the one being served first. The stochastic lead time for a certain order is thereby dependent on previously made orders; however, it is not affected by orders made after the considered order. When calculating the demand during the stochastic lead time it is often practical to use a simple approximation. The requirements are that the demand per time unit is independent and non overlapping. Non overlapping means that orders can't pass each other during the transportation. If the lead time mean and variance is known the mean and variance of the demand during the lead time can then be calculated according to (30) and (31). (Axsäter, 2006)

$$E(D) = \mu E(L) \qquad (30)$$

$$Var(D) = \sigma^2 E(L) + \mu^2 Var(L) \qquad (31)$$

For independent lead times, orders can cross each other in time, which reflects situations where orders are processed by multiple independent servers. In a system with independent stochastic lead times, it's possible for a specific order with a longer lead time to be delivered after another order that was placed earlier.

4.3 Service levels

Another cornerstone in the modeling process is to understand the service levels. Companies frequently utilize service levels as a key performance indicator (KPI) to assess their performance when implementing inventory policies. It is common practice to establish service level constraints to determine the suitable reorder point.

4.3.1 Definition of service levels

Axsäter (2006) defines three different types of service levels denoted S_1, S_2 and S_3 which are stated in (32) , (33) and (34).

$$S_1 = probability\ of\ no\ stockout\ per\ order\ cycle \quad (32)$$

$$S_2 = "fill\ rate" - fraction\ of\ demand\ that\ can\ be\ satisfied\ immediately\ from\ stock\ on\ hand \quad (33)$$

$$S_3 = "ready\ rate" - fraction\ of\ time\ with\ positive\ stock\ on\ hand \quad (34)$$

The first definition, S_1, is easy to use but poses challenges primarily because it neglects batch sizes, making it potentially challenging for analyzing service levels. If S_1 is low but the batch size is big, there can still be a large available stock. Because of this shortcoming this definition of the service levels is not recommended to use in inventory control in practice (Axsäter, 2006).

The fill rate, S_2, and the ready rate, S_3 definitions offer a more comprehensive understanding of the service level, but their computations can be intricate. In scenarios where demand is continuous or follows a pure Poisson distribution, the fill rate and ready rate are identical. However if a customer can order several units at the same time they will no longer be identical. The reason is that it is not sure that the stock will be adequate to fulfill demand even if the existing stock is positive.

In practice using the same service level definition throughout the company which can be followed up using real data is very important in decision making (Axsäter, 2006). Taking a practical point of view it is normally not suitable to set a desired service level for all Stock Keeping Units (SKUs) separately. Instead it is common that SKUs with similar characteristics are segmented into groups on product level (Axsäter, 2006).

4.3.2 Optimizing continuous (R,Q) - policies

Given that a minimum service level is set for a given SKU this can be used as a constraint when optimizing the (R,Q)-policy according to (35) excluding the shortage cost. This approach is commonly used since the shortage cost is more difficult to estimate compared to a service level in general (Axsäter, 2006).

$$min_{R,Q} C(R, Q) = min_{R,Q} (h(IL)^{+} \quad (35)$$
$$s.t \quad SL \geq SL_{Target}$$

Considering a continuous review (R,Q)-policy with a given batch quantity Q it is possible to determine the R that satisfies the given service level for different types of demands. In this section only the fill rate, S_2, and the ready rate, S_3, will be considered.

4.3.2.1 Compound Poisson demand

For compound Poisson demand the probability that the inventory is equal to j is formulated in equation (36). This could be done knowing that the demand during the lead time for a compound poisson distribution follows (13) and the maximum inventory level is $R + Q$ in steady state discussed in section 4.1.3.

$$P(IL = j) = \frac{1}{Q} \sum_{k=max\{R+1, j\}}^{R+Q} P(D(L) = k - j) \qquad j \leq R + Q \quad (36)$$

By definition the ready rate is the probability that the inventory level is positive, see (37).

$$S_3 = P(IL > 0) \quad (37)$$

However, for compound Poisson demand the ready rate and the fill rate are not the same. Calculating the fill rate becomes more challenging due to the variable customer demand quantity. Consequently, the fill rate is determined by considering the expected satisfied quantity for a customer in relation to the expected total demand quantity for a customer, as indicated in equation

(38). Recall that f_k is the probability of demand size k. The fill rate is calculated as the ratio between the expected satisfied quantity and the expected total demand quantity. If the system has a positive inventory level of j and experiences a positive demand size of k units, the delivered quantity is either k units if stock is sufficient, or j units if k > j.

$$S_2 = \frac{E[Satisfied\ Quantity]}{E[Demanded\ Quantity]} = \frac{\sum_{k=1}^{\infty}\sum_{j=1}^{\infty} min(j,k)\cdot f_k \cdot P(IL=j)}{\sum_{k=1}^{\infty} kf_k} \quad (38)$$

4.3.2.2 Normally distributed demand

When the system faces continuous demand that follows a normal distribution, it becomes clear that the fill rate and ready rate are identical. These are calculated as the probability of positive stock on hand according to (39). (Axsäter, 2006)

$$S_2 = S_3 = 1 - P(IL \leq 0) = 1 - F(0) = 1 - \frac{σ'}{Q}\left[G(\frac{R-μ'}{σ'}) - G(\frac{R+Q-μ'}{σ'})\right] \quad (39)$$

In (39), F(x) represents the cumulative cdf for the inventory level x, while G(x) refers to the loss function, defined in accordance with (40) and (41), respectively. In (41), φ and Φ refers to the density and cdf of the standard normal distribution, with a mean of zero and a standard deviation of one, $N(0, 1)$ which are calculated according to (42) and (43), respectively.

$$P(IL \leq 0) = \frac{σ'}{Q}\left[G(\frac{R-μ'}{σ'}) - G(\frac{R+Q-μ'}{σ'})\right] \quad (40)$$

$$G(x) = \int_x^{\infty}(v - x)φ(v)dv = φ - x(1 - Φ(x)) \quad (41)$$

$$φ(x) = \frac{1}{\sqrt{2π}} exp(-\frac{x^2}{2}) \quad (42)$$

$$\Phi(x) = \int_{-x}^{\infty} u\varphi(u)du \qquad (43)$$

4.3.3 Shortage cost

As previously mentioned it is usually perceived as simpler to determine the appropriate service level than estimating a shortage cost (b_1). If the fill rate or ready rate is given it is possible to translate this into an associated shortage cost. This is done according to (44) for Compound Poisson demand and for Normal demand (45). (Axsäter, 2006)

$$S_3(R^*) \leq \frac{b_1}{h+b_1} \leq S_3(R^* + 1) \qquad (44)$$

$$S_2(R^*) = S_3(R^*) = \frac{b_1}{h+b_1} \qquad (45)$$

Here R^* denotes the optimal reorder point. In case of pure Poisson demand $S_2(R^*) = S_3(R^*)$ and (44) is also true for the fill rate (Axsäter, 2006).

4.3.4 Finding the optimal policy parameters with EOQ

When optimizing the (R,Q)-policy discussed in section 4.1.4.1 there are two different possible approaches. The function can be optimized by adjusting the R and Q simultaneously or fixing one of the parameters and then the other. For the second option the approach does not ensure a global optimum however it is a computationally more efficient option.

Mentioned for the first time in 1913 by Ford Whitman Harris, the EOQ-formula has been used for more than a hundred years by decision makers (Harris, 1913). In this formula the economic order quantity is decided according to (46).

$$Q_{EOQ} = \sqrt{\frac{2 \cdot \mu \cdot A}{h}} \qquad (46)$$

Here A denotes the fixed cost associated with planning or producing per order. The EOQ model assumes deterministic constant demand, zero lead time, and that the whole batch is delivered at the same time together with constant ordering and holding cost. With the Q decided the optimal reorder point can be determined according to section 4.1.4.1.

4.4 Inventory control modeling

When modeling an inventory control system the distribution network needs to be constructed. So far in this chapter the focus has been on a single installation, referred to as a single-echelon system. In a single echelon model each installation is modeled in isolation according to Figure 7.

Figure 7: A single echelon inventory system.

However in reality most inventory systems do not consist of independent single echelon systems. In reality inventory systems are often complex consisting of several different installations which are dependent on each other. Stock is transferred between different stocking locations throughout the organization before reaching the final destination, the end customer. A system with multiple stages is called a multi-echelon system.

A multi-echelon system can be constructed in various ways, making it challenging to visualize without additional information. To address this issue, it is beneficial to specify the number of stages and describe the characteristics of the system. Here, the number of vertical stages is emphasized instead of the prefix multi. In Figures 8 and 9, two different types of multi-echelon systems are presented. However, even with this information, understanding the inventory system completely can be difficult. The system in Figure 8 can also be called an One Warehouse Multiple Retailers (OWMR) system.

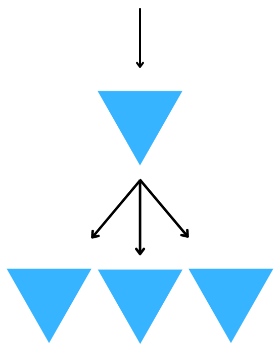

Figure 8: A divergent 2-echelon system with one-warehouse and three dealers.

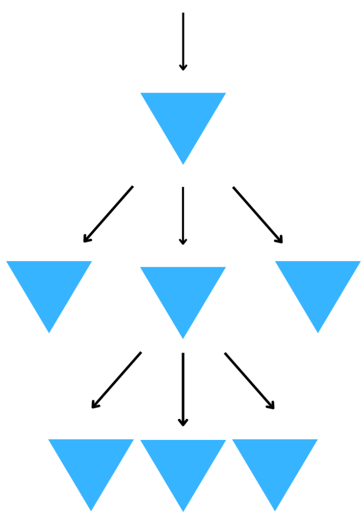

Figure 9: A divergent 3-echelon system.

Two systems that may differ significantly physically can be very similar from an inventory control perspective. For instance, Figure 8 could represent a distribution system where a central warehouse supplies goods to three regional warehouses. Alternatively, it could depict stock positions within a production facility where a subcomponent, used in many products, is stocked at one location, and final goods containing this subcomponent are stored at three different positions.

4.5 Optimization of inventory control systems

Multi-echelon inventory optimization presents tough challenges due to the inherent complexity of real-world systems. Figures 8 and 9 depict two types of multi-echelon systems. In practice, the number of nodes and echelons increase, rendering very complex stochastic systems. In the literature of multi-echelon inventory optimization, there exist exact methods to derive optimal parameters to meet service levels and minimize the expected costs of the system. While exact models offer precision by not incorporating mathematical approximations, they often prove computationally burdensome and applicable only for small systems with specific structures.

The determination of optimal inventory policies for exact models encounters intractability even within rudimentary multi-echelon structures. Despite their ability to mathematically represent the system with high detail, exact models become increasingly cumbersome as system complexity mounts.

In response to these challenges, researchers have embraced approximation techniques and heuristics to facilitate practical implementation. One approach involves decomposing the multi-echelon system into several single-echelon systems, which can then be optimized individually, as discussed in sections 4.1.4.1 and 4.3.2. The objective with these simplified optimization problems is to look for a near-optimal rather than optimal solution for the original multi-echelon problem. When conducting this optimization, it is crucial to appropriately capture the dynamics of the entire system.

Further complexity arises in the optimization problems when considering additional factors such as service differentiation. The ability to provide customers at the same stock point with different service levels is one common requirement by industry. These additional complexities contribute to the challenge of finding optimal inventory policies. In the following sections service differentiation is discussed further.

4.5.1 Reasons for service differentiation

Service differentiation in inventory management involves customizing inventory policies and service levels to accommodate distinct requirements and preferences of different customer segments, distribution channels or product categories. Segmentation is crucial in inventory control, this is illustrated in the following scenarios:

The criticality of spare parts may vary across different locations or plants. While a part could be deemed critical in one facility, it might be considered non-critical in another. In instances where backorders occur, the costs associated with critical parts are typically significantly higher than those of non-critical parts. (Dekker et al., 1998)

Parts may be ordered for regular restocking purposes or due to emergency requirements. Emergency orders often entail higher costs, especially when utilizing expensive transportation modes such as airline distribution. Consequently, the cost of backorders for emergency orders tends to be higher compared to regular orders. (Nahmias and Demmy, 1981)

Certain customers may hold greater importance to the business than others. Therefore, prioritization within inventory management based on customer importance ensures that critical customers receive adequate attention and service levels. (Shulte and Pibernik 2017) (Teunter and Hanevald 2007)

There is a growing trend where customers are offered various service level agreement contracts to choose from. This trend emphasizes the need for companies to tailor their inventory management practices to meet the diverse service expectations of different customer segments. (Shulte and Pibernik 2017) (Teunter and Hanevald 2007)

4.5.2 Service differentiation modeling

Acknowledging and effectively managing the intricacies of segmentation in inventory management empowers businesses to streamline resource allocation, reduce expenses and elevate customer satisfaction. While it is evident that a priority clearing mechanism, prioritizing backorders for high-priority customers is practical, implementing such a policy poses analytical and tractability challenges, as noted in Berling et al. (2023).

In research on service differentiation within inventory management, considerable attention is devoted to inventory rationing, also known as stock reservation. A majority of academic journals identified through a search for "service differentiation AND inventory management/control" on LubSearch focus on this approach. Further searches using "stock rationing/inventory rationing AND inventory management/control" yield numerous additional research articles. Among the proposed strategies, a significant portion emphasizes the critical-level policy, first introduced by (Veinott, 1965), which involves allocating a buffer stock to meet the demands of customers with higher target service levels.

Research on inventory rationing is broadly categorized into two main approaches: cost-based and service level-based. The cost-based approach focuses on minimizing expenses associated with inventory management, while the service-level approach aims to meet or exceed customer service objectives. These two approaches can further be delineated into four dimensions.

The first dimension involves static versus dynamic rationing. Dynamic rationing methods are notably more computationally complex, considering factors such as the expected remaining time until the next inventory review or the remaining lead time until the next order arrives. The second dimension pertains to the number of service classes, which can be classified into two options: either two classes or an arbitrary number of classes. This dimension correlates directly with the number of distinct service levels. The third dimension considers periodicity, distinguishing between single-period and multiple-period inventory rationing. In single-period inventory rationing, decisions are made for a single time period, often representing a single outstanding order. Conversely, multiple-period inventory rationing spans over multiple periods, accommodating several outstanding orders. The fourth dimension addresses the treatment of shortages as either backordered or modeled as lost sales. (Alvarez et al., 2013)

4.5.2.1 Service differentiation in single-echelon models

In their comprehensive literature review, Teunter and Hanevald (2007) delve into multiple models proposed by researchers aimed at integrating service differentiation into stock rationing strategies. A recurring theme across these models is the notable computational complexity and time-intensive optimization processes. Addressing these challenges, they present a single-period model designed to establish a critical order policy accommodating two demand classes (critical and non-critical), operating under the assumption of Poisson demand and backordering. Conversely, Schulte and Pibernik introduce another model capable of handling an arbitrary number of classes. They develop a closed-form expression to determine service levels for N service classes, an advancement previously limited to scenarios with only two classes (Schulte and Pibernik 2016).

Similar to the two models presented above, under assumption of poisson demand, Arlsan et al. (2007) have developed a model for an arbitrary number of service classes with a clearing mechanism for backorders, which treats a backorder from a lower-priority class equivalent to a first come first served (FCFS) policy. This means that when the inventory position falls below certain thresholds, only higher priority demand gets served. The aggregate inventory stock for each service class can be represented by a serial system, illustrated in Figure 10, wherein the various service classes are divided across N installations. This approach effectively divides the on-hand inventory of different demand classes into N distinct stockpiles, where i = 1 has the highest service level requirement. (Arslan et al., 2007)

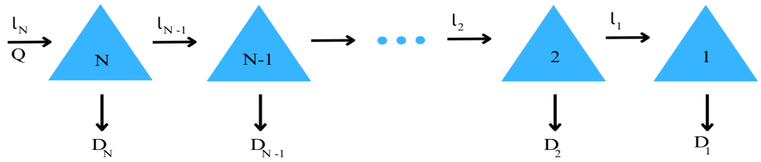

Figure 10: A serial inventory system with demand at each installation.

Moreover, each installation adopts a continuous review base-stock policy with Poisson distributed demand, a nonnegative base-stock level s_i, and is replenished directly by the preceding installation with zero lead time. The base-stock level s_i is derived as the difference between the critical level, c_i, and the preceding critical level c_{i-1}, where $c_{i-1} \leq c_i$ and $c_0 = 0$. This base-stock level s_i serves as a reservation stock for class *i*, representing the inventory available to fulfill class-i demand and replenish the reserve stocks for any higher-priority demand classes. Stage N replenishes from an outside supplier with lead time greater than zero, and uses an (R,Q) policy. Given the system set up, the reorder point R at stage N can be according to (47).

$$R = \sum_{i=1}^{N} s_i \qquad (47)$$

Thus, R can be interpreted as the reorder point of the total inventory to serve the different service classes. They later present an exact method for determining optimal parameters which yields promising results. They also suggest how the model can be extended to a multi-echelon setting and handle general demand distributions, but refer to this as future studies. (Arslan et al., 2007)

4.5.2.2 Service differentiation in multi-echelon models

Transitioning to the exploration of service differentiation within multi-echelon environments, Axsäter et al. (2004) introduced a tailored multi-echelon service differentiation model specifically designed for an One-Warehouse-Multiple-Retailers (OWMR) system. Their model places significant emphasis on managing backorders resulting from direct demand at the central warehouse. Central to their framework is the establishment of critical stock levels for each retailer, defining the threshold at which a replenishment order is fulfilled from the warehouse inventory. This means that a retailer's order is processed from the warehouse stock only when the stock level exceeds this critical threshold. Additionally, it is assumed that if a retailer's replenishment order cannot be fulfilled from the warehouse stock, it can always be sourced, as an emergency order, at a higher cost from an external supplier, rather than being backlogged. (Axsäter et al., 2004)

Although the results are promising, the model's applicability is restricted by its fundamental assumptions. Limitations such as the uniform adoption of an (S-1, S) inventory policy across all locations, the assumption of Poisson-distributed demand, and the absence of backlogging for retailer orders impede its broader applicability.

In a subsequent article, Axsäter et al. (2007) presents another method investigating service differentiation within a two-echelon distribution system. Departing from the utilization of critical policies, they advocate for a strategy involving the reservation of separate stock to manage direct upstream demand at the central warehouse (Axsäter et al., 2007). This method introduces an artificial retailer at the central warehouse with zero transportation time, offering an attractive approach due to its compatibility with established models for OWMR-systems. However, as highlighted in (Berling et al., 2023), the performance of this separate stock approach deteriorates in systems characterized by fill rate constraints and significant variations in customer order sizes. In such scenarios, the order-up-to levels at the artificial retailer are often overestimated when treated in the same way as regular retailers with a positive lead-time.

4.5.2.2.1 Controlling inventories in omni/multi-channel distribution systems with variable customer order-sizes

Driven by gaps in the research literature, and industry collaboration, (Berling et al., 2023) have devised a combined stock approach to manage inventory systems in one-warehouse-multiple-retailer setups, where the central warehouse faces direct customer demand. This combined stock method for service differentiation at the central warehouse, can be conceptualized as a critical level policy. Their computationally efficient heuristics address the complexities of real-world one warehouse multiple retailer inventory systems, which are characterized by highly variable customer order sizes, (R, Q) policies implemented across all stock points, and fill rate constraints. All unsatisfied demand is backordered and satisfied in a first come first served (FCFS) manner.

Similar to (Axsäter et al., 2007), they introduce an artificial retailer at the central warehouse to serve direct upstream demand. The stock at the central warehouse can then be divided into two parts: reservation stock dedicated to serve the direct upstream demand, and the general warehouse stock which is used to replenish all retailers including the artificial retailer. The artificial retailer replenishes from the general stock according to a continuous review (S-1, S) policy, where the order up to level, S, can be interpreted as the critical reservation level for the combined stock at the central warehouse. A representation of the system can be seen in Figure 11.

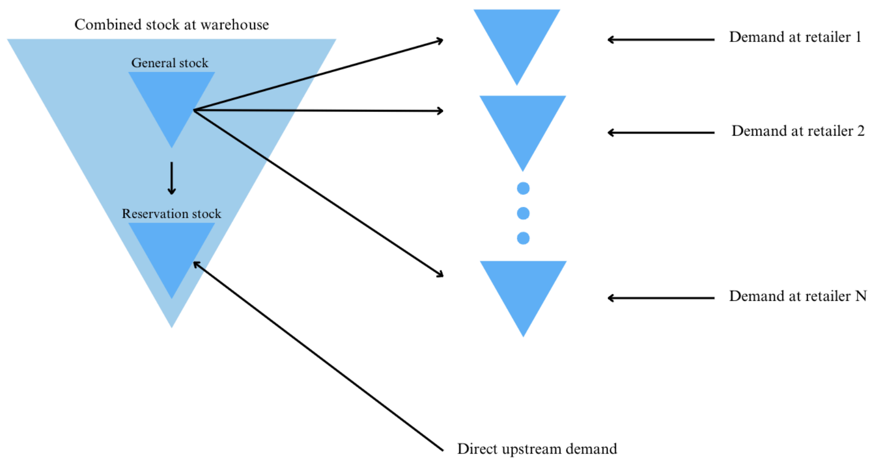

Figure 11: A representation of a system with general and reservation stock.

The combined stock heuristic takes into account the entirety of available inventory within the central warehouse when establishing the critical reservation level, thereby not overestimating the order-up-to level S at the artificial retailer. This heuristic is structured into two distinct steps. Initially, the focus is on determining the critical reservation level. The goal here is to identify the smallest S value that fulfills the fill rate constraint for direct upstream demand, given a reorder point R_0 for the general warehouse stock. Subsequently, the second step involves integrating the combined stock heuristic into an efficient inventory control method for the entire OWMR system. This entails optimizing the reorder point R_0 and R_i for i = 1, 2... N.

Drawing upon the findings in previous articles (Berling and Marklund 2013) and (Berling and Marklund 2014) the authors propose integrating the combined stock heuristic with the inventory optimization model tailored for scenarios involving compound Poisson and normally distributed demand. A comprehensive numerical study demonstrates that this integrated approach yields near optimal solutions with respect to target fill rates, while simultaneously offering substantial potential for minimizing inventory costs.

Since the article from 2023 was published (Berling et al, 2023), Prof Johan Marklund has continued his research on service differentiation within multi echelon settings. The method proposed above can handle two demand channels, direct demand at the central warehouse, and demand at the central warehouse triggered by retailers. In an unpublished article (work in progress) he has proposed a method for further service differentiation for an arbitrary number of distribution channels, introducing additional direct demand channels at the central warehouse. The model will be explained in the next chapter. For the special case of three distribution channels applicable to Volvo's system, this will be referred to as the EM-model (Extended Model).

Chapter 5. Inventory control of spare parts at Volvo

In the context of Volvo's Spare Parts distribution network and inventory policies described in section 1.3, the existing models in the field quickly reveal their limitations. Volvo operates with a continuous review (R,Q)-policy for inventory management, serving a diverse network of non-identical dealers with varied demand patterns. Much of the research on multi-echelon models assumes centralized control over all installations, a condition not applicable to Volvo's distribution network where independent dealers operate. With reference to the preceding literature review, it becomes evident that a model accommodating a multi-echelon setup while addressing service differentiation across multiple distribution channels without model-restrictive assumptions is essential.

Considering the existing distribution network, the EM-model emerges as the best fit for the Volvo Spare Parts organization's inventory control system. This model aligns with Volvo's structure and requirements. The EM-model accounts for (R,Q) policy and is compatible with demand distributions such as normal and compound Poisson. Independent dealers, whose inventory policies Volvo cannot directly dictate, are modeled as direct upstream demand at the central warehouse. The direct demand stemming from non-controllable retailers can be divided into two separate distribution channels, depending on the service level requirement. As for Volvo-owned and LPA-dealers (i.e. controllable retailers), they will integrate into the multi-echelon system, allowing for the derivation of optimal inventory policies for each dealer.

In the following sections, a conceptual description of the EM-model will be provided, followed by a detailed, step-by-step approach to derive optimal inventory parameters within the system. But before delving into the model description, it is crucial to understand the method for estimating stochastic lead time, as this is essential for grasping the processes within the model. The next section offers the necessary background to comprehend the lead time approximation.

5.1 METRIC inspired approximation of lead times

The so called METRIC approximation stems from the seminal paper by Sherbrooke (1968). This paper studies a system with base-stock policies and Poisson demand at all installations. The idea is to replace the stochastic lead times by their averages, In this approximation, the lead time comprises of a constant transportation time between stocking locations and a stochastic waiting time caused by stock-outs at the warehouse, resulting in delays. This is calculated according to (48).

$$\bar{L}_i = L + E(W_0) \tag{48}$$

In this model, the stochastic waiting time is represented as the mean of the stochastic waiting time, W_0, at the warehouse. This approximation works well also under different model assumptions (Andersson et al., 1998). For Poisson demand and base-stock policies at all the retailers, or if all retailers use the same order quantity, a direct application of Little's law, see (49), renders the correct mean of $L_i(R_0^*)$. Where μ_0 denotes the mean subbatch demand at the central warehouse during L_0. $E[B_0(R_0^*)]$ is the expected number of backorders.

$$\bar{L}_i(R_0^*) = l_i + \frac{L_0}{\mu_0} E[B_0(R_0^*)] \tag{49}$$

In the case of retailers with different order quantities and/or more general compound Poisson demand, (49) does not represent the correct mean. However, the result serves as an approximation (Berling and Marklund 2013).

5.2 The multi-echelon EM-model

The EM-model, as previously mentioned, is a special case of a more general model developed by Prof. Johan Marklund, based on prior work with Peter Berling and Lina Johansson (Berling et al., 2023). This OWMR/omni-channel model features a serial system consisting of two virtual retailers and one distribution channel dedicated to serving demand from controllable retailers.

This omni/multi-echelon One-Warehouse-Multi-Retailer (OWMR) inventory system accommodates end customer demand originating from both the central warehouse (CW) and multiple non-identical retailers. Demand from non-controllable retailers is represented as direct demand to the central warehouse, channeled through either VR_1 or VR_2 based on the prescribed service level requirement. The model manages two distinct service channels for processing direct demand, where demand entering VR_2 has a higher target service level than VR_1. The total stock at the CW is divided into reservation stock, dedicated to serving direct upstream demand (supplying demand entering the virtual retailers), and general warehouse stock, used to replenish all retailers, including the virtual retailers. For a visual representation of the model, see Figure 12.

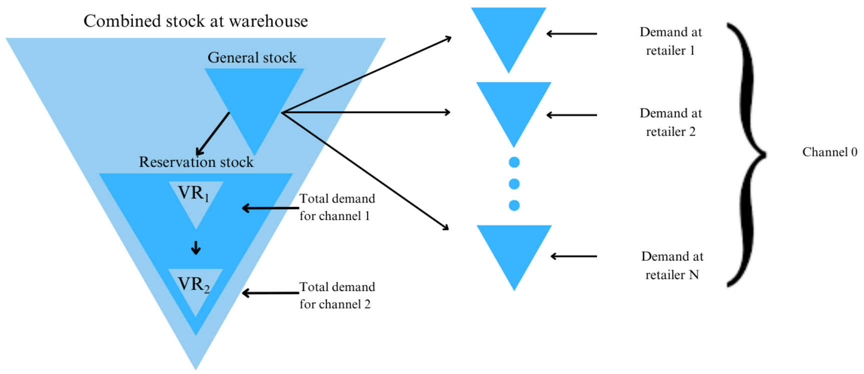

Figure 12. A representation of the EM-model.

Lead time to the CW, L_0, and transportation times from the CW to regular retailers, l_i, are assumed to be positive and constant, which is standard in inventory control modeling. The transportation time may be subject to uncertainties; however, the model can accommodate this by applying standard approximation techniques, as briefed in section 5.1, for dealing with stochastic lead times, replacing the stochastic lead time with its mean.

In the model, transportation times between the CW and virtual retailers are zero since virtual retailers model the reservation stock at the CW, which is part of the total CW stock. Despite the zero transportation time, there may be a positive lead time to the virtual retailers caused by stockouts. Similarly, for regular retailers, the lead times are stochastic due to stockout probabilities at the CW.

All inventory locations are assumed to apply continuous review installation stock (R,nQ) policies to replenish inventory. Thus, an order size of Q is placed when the inventory position falls to or below the reorder point R. The virtual retailers are integrated with the CW and are replenished from the general stock according to a continuous review (S−1, S) policy, equivalently an (R, 1) policy. The order-up-to levels (R+1), in this model denoted S_1 and S_2 for each virtual retailer, can be interpreted as the critical reservation levels for the combined stock at the CW.

All unsatisfied demand is backordered and satisfied on a first-come-first-serve (FCFS) basis. This means there is no difference in priority between regular and virtual retailers when clearing backorders. This is a common assumption in the literature and used for example in Arslan et al. (2007), Axsäter et al. (2007) and Berling et al. (2023).

The considered costs include holding costs per unit and time unit at each inventory location. All retailers, including the virtual retailers, operate under fill rate constraints with specified fill rate targets, γ_i^*. The fill rate is defined as the fraction of demand that can be satisfied directly from stock on hand.

The objective is to minimize the expected holding cost per time unit, TC, by optimizing R_0 for the general stock, the critical reservation levels, S_1 and S_2 for the virtual retailers, and R_i for the regular retailers, subject to fill rate constraints. This can be modulated according to (50).

$$\min TC(R_0, S_1, S_2, R_i, \ldots, R_N) = h_0 E[IL_{CW}^+] + \sum_{i=1}^{N} h_i E[IL_i^+] \quad (50)$$

$$s.t.\ \gamma_{VR(i)}(R_0, S_i) \geq \gamma_{VR(i)}^*$$

$$s.t.\ \gamma_i(R_0, R_i, \ldots, R_N) \geq \gamma_i^*$$

In order to solve the reservation levels for the virtual retailer, the combined stock on hand, i.e. reservation stock and the general stock, need to be considered since both sources can supply the virtual retailers. In the next section, the combined stock heuristic originating from Berling et al. (2023) is presented.

5.2.1 The combined stock heuristic

We assume a central warehouse where the total stock is divided into two parts: a general stock supplying both regular retailers and virtual retailers and a separate stock only satisfying demand from virtual retailers. The aim is to obtain a base stock level for VR_1 and VR_2 such that the combined stock at the CW and the virtual retailers, together achieve the target fill rates in the three distribution channels. The reorder point for the combined stock R_{CW} is defined according to (51).

$$R_{CW} = R_0 + \sum_{j=1}^{2} S_j \quad (51)$$

The base stock level S_j can be interpreted as the critical level or reservation stock associated with channels $i \leq j$. When the combined stock on hand drops below this threshold, only demand from the associated channels are supplied. If $j = 1$, customers from channel VR_1 and VR_2 are supplied. If $j = 2$, only customers in channel VR_2 are supplied, see Figure 13.

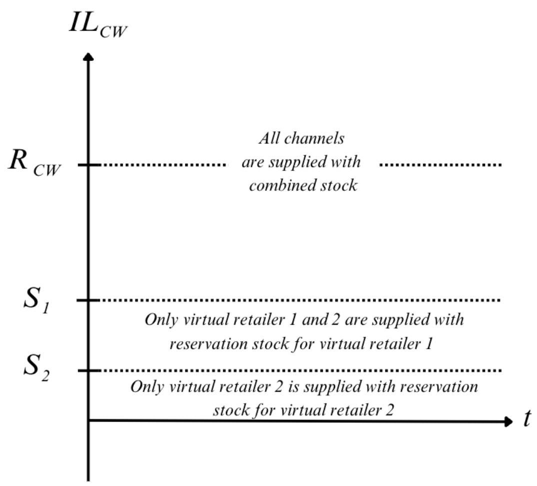

Figure 13: Illustration of the reservation stock with two virtual retailers.

The approach is based on approximating the probability mass function (pmf) for the combined inventory level of the general stock and the reservation stock at the central warehouse, given the reservation levels for the virtual retailers and the reorder point at the

central warehouse, R_0. The combined stock policy prescribes that when a customer arrives and demands d units, it will be satisfied by the combined stock on hand at the central warehouse IL^+_{CW}. IL^+_{CW} is the sum of the available stock on hand at the artificial retailers ($\leq S_1+S_2$) and the available general warehouse stock. Thus, even if $d > S_1 + S_2$ the demand may be fully satisfied if $d - S_1 - S_2$ units are available in the general warehouse stock.

The challenge using the combined stock is to determine the probability distribution of the combined stock on hand IL^+_{CW}. The exact distribution is unknown and inherently complex, and for computational reasons most likely infeasible to use in the real system. Therefore, the focus is on obtaining an efficient approximation.

The approach and determination of the reservation levels for the virtual retailers are determined by sequential computation. First, the reservation level for VR_1 is computed. Essentially, the inventory level distribution of the general stock and reservation stock for the first virtual retailer, $IL_{0,1}^{tot}$, is determined. Given the distribution of the inventory level, the reservation level can be computed. S_1 should be set as low as possible while still fulfilling the target fill rate, γ_1^*. It is straightforward to determine the smallest S_1 that satisfies the target fill rate by increasing S_1 from zero until $\gamma_1(R_0, S_1) \geq \gamma_1^*$. When the reservation level for VR_1 is set, the reservation level for VR_2 can be computed in a similar manner.

The probability distributions of the inventory levels for the virtual retailers are difficult to determine, as the replenishment lead-time is stochastic because the general stock may be depleted. The approximation for this stochasticity is determining a mean value of the delay to virtual retailers caused by stock outs at the general stock. Focusing only on these situations, i.e. when $IL_0 \leq 0$, an estimate of the inventory distribution can be determined, and the reservation levels can be calculated.

It is worth mentioning that the method of determining the inventory level for the combined stock assumes independence between the inventory level for the general stock, IL_0, and the inventory levels for the virtual retailers. Even if this is not the case, it is assumed as an approximation for the problem to be computationally feasible.

5.3 Solving optimal inventory policies using the EM-model

The overall approach for obtaining optimal policies using the EM-model can be divided in four steps. The first step revolves around determining the lead time demand at the central warehouse, setting the foundation for determining the optimal reorder point in the subsequent step. The second step is to determine a near optimal reorder point for the general stock at the central warehouse, R_0. This general stock is used for serving all channels, but it is also the only stock used for replenishing the regular retailers in the original OWMR system. In the chosen approximation model based on (Berling and Marklund 2013, 2014), this means estimating near optimal induced backorder costs for the different channels. The third step concerns finding optimal reorder points for the regular (Volvo owned and LPA) retailers. Lastly, the fourth step is concerned with determining the base-stock levels at the virtual retailers (supplying independent retailers) to achieve the specified target fill rates for the virtual retailers with successively higher target fill rates.

The solution is essentially broken down into a series of single-echelon problems. The decomposition of the system is achieved by introducing an induced backorder cost, β, at the central warehouse, that captures how the retailers are affected by the reorder point R_0 for the general stock at the central warehouse. Once the optimal reorder point at the central warehouse is set, the reservation levels for virtual retailers and regular retailers can be optimized independently to meet the specified service level targets.

Step 1: The distribution of the lead-time demand at the central warehouse, $D_0(L_0)$, is determined by treating the virtual retailers as any other retailer. $D_0(L_0)$ incorporates the impact of retailer order quantities and customer order-sizes.

Step 2: In order to determine the reorder point for the general stock, the induced backorder cost for the central warehouse is calculated. The induced backorder cost is estimated as a demand weighted average of the induced backorder costs associated with each retailer. In our case, this includes both the regular and virtual retailers. The induced backorder cost should capture the effect of how changes in lead-time (i.e. a new reorder point at the central warehouse) impact the retailers. When the induced backorder costs have been estimated, a near optimal reorder point at the general warehouse stock, R_0, is calculated. That is, by minimizing the expected holding and induced backorder cost per time unit at the central warehouse, according to (52).

$$minC_0(R_0) = \frac{h_0+\beta}{Q_0} \sum_{y=R_0+1}^{R_0+Q_0} E_{D_0(L_0)}[(y - D_0(L_0))^+] - \beta Q(R_0 + \frac{Q_0+1}{2} - \mu_0)$$

(52)

It is easy to show that $C_0(R_0)$ is convex in R_0 and that the optimal reorder point R_0^* that minimizes $C_0(R_0)$ can be found through a simple search using the optimality condition according to (53).

$$R_0^* = max\{R_0: C_0(R_0) - C_0(R_0 - 1) \le 0\} \tag{53}$$

Step 3: A near optimal reorder point at each regular retailer is determined by solving the fill rate constrained singe-echelon problem according to (54), where $D_i(\bar{L}_i)$ is the demand at retailer i during time period $\bar{L}_i$.

$$min\ C_i(R_i) = \frac{h_i}{Q_i} \sum_{y = R_i+1}^{R_i+Q_i} E_{D_i(\bar{L}_i)}[(y - D_i(\bar{L}_i))^+] \tag{54}$$
$$s.t.\ \ \gamma_i(R_0, R_i) \ge \gamma_i^*$$

Step 4: The critical reservation levels for the virtual retailers is determined using the combined stock heuristic. The inventory level distribution of the combined stock is calculated and used for determining the critical reservation levels necessary to achieve the target fill rates for the virtual retailers.

5.3.1 Step 1. Lead time demand at the central warehouse

The lead time demand at the central warehouse, $D_0(L_0)$ is expressed in "subbatch" demand rather than in unit demand. The batch quantity at retailer i, q_i, is expressed as the subbatch quantities of Q, where Q is the largest common divisor of all order quantities. Because the batch quantity for all virtual retailers equals one, the largest common divisor of all order quantities equals one in the system, and the "subbatch" demand can be interpreted as unit demand.

The demand from retailers is assumed to follow a compound poisson distribution, see (11)-(13). Furthermore, an assumption of the EM-model is that the lead time for an order to arrive at the central warehouse from an outside supplier, L_0, is constant. The derivation of the lead-time demand can then be determined according to the process below.

Firsty, the probability distribution of the lead time demand associated for each retailer need to be computed. This is done according to equation (55), where $\delta_i(n)$ is the probability that at most *n* orders from retailer *i* have been placed during L_0 time units. For the specified retailer model, the inventory position at retailer i, IP_i, is uniformly distributed over $[R_i + 1, R_i + Q_i]$ as long as all demands are not multiples of an integer greater than one. Thus, if we consider retailer i at an arbitrary time and let $x = IP_i - R_i$, it follows that x is uniformly distributed over $[1, Q_i]$.

$$\delta_i(n) = \frac{1}{Q_i}\sum_{x=1}^{Q_i} P(D_i(L_0) \le nQ_i + x - 1) \quad for\, n = 1, 2, 3... \tag{55}$$

Defining $D_o^i(L_0)$ as the subbatch demand from retailer *i* during L_0 time units, with a probability mass function (pmf) $g_0^i(u)$, the subbatch demand for retailer *i* can be determined according to (56).

$$g_0^i(u) = P\left(D_0^i(L_0) = u\right) = \begin{cases} \delta_i(0) & \text{if } u = 0 \\ \delta_i(n) - \delta_i(n-1) & \text{if } u = nq_i,\ n = 1,2,... \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

Once $g_0^i(u)$ is determined for every retailer, it is straightforward to calculate the mean and standard deviation of the lead-time demand at the central warehouse according to (57) and (58).

$$\mu_0 = \mu_0^1 + \mu_0^2 + ... + \mu_0^{N+M},\ where\ \mu_0^i = \frac{\mu_i L_0}{Q} \tag{57}$$

$$\sigma_0^2 = (\sigma_0^1)^2 + (\sigma_0^2)^2 + \ldots + (\sigma_0^{N+M})^2 \ where\ (\sigma_0^i)^2 = \sum_{u=0}^{\infty} (\mu_0^i - u)^2 \cdot g_0^i(u)$$
(58)

Knowing the mean and standard deviation for the demand during the lead time at the central warehouse, this can be used to fit either a normal distribution, gamma distribution or a negative binomial distribution to approximate the lead time demand.

When $\frac{\sigma_0^2}{\mu_0} \geq 1$ the negative binomial distribution is used, if $\frac{\sigma_0}{\mu_0} < 0{,}2$, the normal distribution is used, and if neither of the two conditions are met, the gamma distribution is used.

For the negative binomial distribution, the parameters r and p can be determined according to (18) and (21). Given r and p, the demand during the lead-time L_0 can be determined according to (59).

$$P\left(D_0\left(L_0\right) = u\right) = g_0(u) \cong \begin{cases} (1-p)^r & for\ u = 0 \\ \frac{r(r+1)\ldots(r+u-1)}{u!} \cdot (1-p)^r \cdot p^u & for\ u = 1,2,\ldots \end{cases}$$
(59)

If the normal distribution or gamma distribution is used, the demand during the lead time is determined according to (60), where F(x) is the cumulative distribution function (CDF) of the distributions.

$$P\left(D_0\left(L_0\right) = u\right) = g_0(u) \cong \begin{cases} F(0{,}5) & for\ u = 0 \\ F(u+0{,}5) - F(u-0{,}5) & for\ u = 1,2,\ldots \end{cases}$$
(60)

When the central warehouse demand distribution is determined, we can proceed to determine the induced backorder cost.

5.3.2 Step 2. Determine the induced backorder costs and calculate R_0

In this section, we first introduce the concept of induced backorder cost to provide a foundational understanding of this critical step. Next, we present two distinct approaches for determining the induced backorder cost at virtual retailers. Finally, we delve into the iterative procedure used in our analytical model for calculating the induced backorder cost for virtual retailers, offering a detailed explanation of the process. In the process of finding estimates of the induced backorder cost for the virtual retailers, the optimal R_0 is determined. Readers who require no background are referred to section 5.3.2.4 for the algorithmic derivation of the induced backorder cost and the reorder point at the central warehouse.

5.3.2.1 Induced backorder cost

To enhance the efficiency of the decomposition model described in Andersson et al. (1998), Berling and Marklund (2006) provide a closed-form estimate of the optimal penalty cost, referred to as the induced backorder cost, denoted β_i. This cost reflects the impact of the reorder point R_0 has on retailers and is derived as a scaled derivative of the optimal retailer costs with respect to the average lead time (Berling and Marklund, 2006). The closed form expression of the induced backorder cost β_i is determined according to (61). See Appendix C for all equations to determine β_i.

$$\beta_i = h_i \cdot g(Q_{i,n}, p_{i,n}) \cdot \sigma_{i,n}^{k(Q_{i,n}, p_{i,n})}, \quad For\ i = 1, 2 \ldots N \tag{61}$$

The calculations for the induced backorder costs are based on a normalized system where a unit of demand equals 100, the holding cost per unit and time unit equals 1, and the transport time from warehouse to retailer equals 1 time unit. Parameters from any system can be scaled to this normalized model using the conversion formulas provided in the Appendix C. The system needs to be normalized before one can use the closed form expressions.

The induced backorder costs may vary among non-identical retailers. Therefore, the induced backorder cost at the central warehouse is calculated using a weighting scheme. Berling and Marklund (2006) examined several weighting schemes and found no significant differences. Consequently, in their later works (Berling and Marklund, 2013; 2014), they adopted a simple weighting scheme based on the proportion of total expected customer demand , see (62).

$$\beta^* = \sum_{i=1}^{N} \frac{\mu_i \cdot L_o}{\mu_0 \cdot Q} \cdot \beta_i \qquad (62)$$

5.3.2.2 Two approaches for estimating the induced backorder cost for virtual retailers

Since the virtual retailers in the model are integrated in the central warehouse, the transportation time is zero. Therefore, the closed form induced backorder cost approximation presented by Berling and Marklund (2006) cannot be used for determining the induced backorder cost for virtual retailers. To tackle this Berling et al., (2023) proposes a naïve approximation where the induced backorder cost, β_{VR}, is equal to the backorder cost associated with the direct customer demand, $\hat{p}$. Here it is proposed that this can be estimated according to (63) if not provided as an input variable. In this estimate the target service level, FR_{VR}, and the holding cost, h, is expected to be known.

$$FR_{VR} = \frac{p}{p+h} \Rightarrow \hat{p} = \frac{FR_{VR} \cdot h}{1 - FR_{VR}} \qquad (63)$$

The naïve approximation tends to overestimate the optimal induced backorder cost. This leads to higher inventory levels at the central warehouse due to setting the reorder point at the central warehouse too high. (Berling et al., 2023)

An alternative approach involves utilizing an iterative procedure for establishing the induced backorder cost, as outlined in the paper presented by Andersson et al. (1998). This method typically yields a lower value for the induced backorder cost at the virtual retailers (Berling et al., 2023) resulting in lower inventory at the central warehouse. While the original paper was derived for normally distributed demand, Berling et al. (2023) suggest and evaluate its application also for compound Poisson distributed demand.

5.3.2.3 Iterative procedure for determining the induced backorder cost and R_0

Since this is a computational heavy method, this section will first provide a conceptual description of the iterative procedure. The later part provides a stepwise approach in order to compute the induced backorder costs and an optimal R_0 at the central warehouse.

The method iteratively updates the induced backorder costs for VR_1 and VR_2 for new values of R_0 until an equilibrium is reached. Throughout this iterative procedure, the estimates of the induced backorder costs for VR_1 and VR_2 are continuously updated, feeding back into recalculating the reorder point at the central warehouse. The cycle of adjustments and repeats until the new set of calculated induced backorder costs for VR_1 and VR_2 results in an R_0 equal to the previous iteration, ensuring the reorder point at the central warehouse reaches a stable value. For an illustrative flow chart of the process, see Figure 14.

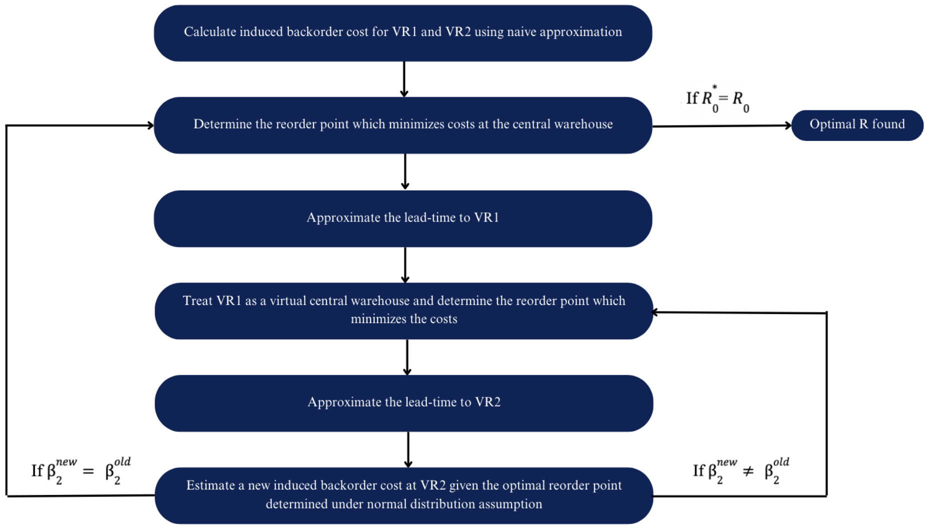

Figure 14: Flow chart of the iterative procedure for determining R_0.

Step 2.1: Calculate the reorder point at the central warehouse based on the naïve approximation of the induced backorder costs for virtual retailers, and use closed form expression for regular retailers.

Step 2.2: Use the expected number of backorders at the central warehouse to calculate the lead time to VR_1.

Step 2.3: Treat VR_1 as an artificial central warehouse supplying direct demand entering VR_1 and indirect demand entering VR_2. Determine the reorder point which minimizes the expected holding and induced backorder costs.

Step 2.4: Given the reorder point for VR_1, determine the expected number of backorders and calculate the expected lead time to VR_2.

Step 2.5: Assume demand at VR_2 is normally distributed and estimate β_{VR2} as the derivative of the optimal expected holding and backorder costs with respect to its lead time. Iteratively return to step 2.3 until equilibrium is reached.

Step 2.6: Use the new estimate of the induced backorder cost for VR_1 and VR_2 to calculate a new reorder point at the central warehouse.

Step 2.7: Proceed from step 2.2 through step 2.6 until the reorder point, R_0, does not change from one iteration to the next, then the procedure stops.

5.3.2.4 Algorithmic derivation of the induced backorder cost and reorder point R_0

In this section, the index j refers to the virtual retailers, while the index i refers to the regular retailers.

Step 2.1.1 Determine the induced backorder costs, β_i for all regular retailers i=1,2,….,N This is done using the closed form estimates presented in section 5.3.2.1, where the system is first normalized and the induced backorder cost calculated.

Step 2.1.2: Determine initial naïve estimates of the induced backorder costs for the virtual retailers. Start from VR_2 and use the naïve estimate $\beta_2 = p_2$. Typically, the target fill rate FR_2 is given rather than the backorder cost p_2. In this case one may use the standard estimate according to (64).

$$FR_{VR} = \frac{p}{p+h} \Rightarrow \hat{p} = \frac{FR_{VR} \cdot h}{1 - FR_{VR}} \tag{64}$$

Determine an estimate for the induced backorder cost for VR_1 as demand weighted averages of the induced backorder cost using (65)-(67).

$$\hat{p}_j = h_0(FR_j) / (1 - FR_j) \tag{65}$$

$$p_j = \beta_{j+1}\left(\frac{\mu_{j+1}^{tot}}{\mu_j^{tot}}\right) + \hat{p}_j\left(\frac{\mu_j}{\mu_j^{tot}}\right) \tag{66}$$

$$\beta_j = p_j \tag{67}$$

step 2.1.3: Optimize R_0 for the given induced backorder costs. Use the weighted induced backorder cost β to optimize the reorder point R_0

according to (68) Note that the induced backorder cost for VR_1 comprises estimates for VR_2. Also note that $\frac{\mu_0 Q}{L_0} = \sum_{i=1}^{N} \mu_i$.

To clarify the expression, j=1 refers to VR_1.

$$\beta = \left(\sum_{i=1}^{N}\left(\frac{\mu_i \cdot L_0}{\mu_0 \cdot Q} \cdot \beta_i\right)\right) + \left(\frac{\mu_{j=1}^{tot} \cdot L_0}{\mu_0 \cdot Q} \cdot \beta_{j=1}\right) \qquad (68)$$

Given the induced backorder cost β, the expected cost of a given reorder point can be computed according to equation (69). Since β is larger than zero and $C_0(R_0)$ - $C_0(R_0 - 1)$ is increasing in R_0 it is easy to show that this implies that $C_0(R_0)$ is convex. Knowing this the optimal reorder point at the central warehouse can be found according to (70). Recall that $g_0(u)$ is determined according to section 5.3.1 and that Q equals one for this model.

$$C_0(R_0) = \frac{(h_0+\beta)\cdot Q}{Q_0}\left\{\sum_{y=R_0+1}^{R_0+Q_0} \cdot \sum_{u=0}^{y}(y-u)g_0(u)\right\} - \beta Q\left(R_0 + \frac{Q_0+1}{2} - \mu_0\right) \qquad (69)$$

$$R_0^* = max\left\{R_0 : C_0(R_0) - C_0(R_0 - 1) \leq 0\right\} \qquad (70)$$

If the naïve estimates are going to be used and no improvement using the iterative procedure is desired the procedure ends here. Otherwise proceed. If $R_0^* = R_0$, the reorder point corresponds to the previous set of induced backorder costs, and no further updating of the induced backorder costs and R_0 is needed. Otherwise if $R_0^* \neq R_0 \Rightarrow$ set $R_0 = R_0^*$ and proceed to **step 2.2.1**.

Step 2.2.1: Estimate the associated expected lead-time, $\overline{L}_{j=1}(R_0)$ for VR_1 according to the METRIC inspired approach in (71) where $E[B_0(R_0)]$ is calculated according to (72).

$$\overline{L}_{j=1}(R_0^*) = \frac{L_0}{\mu_0} \cdot E[B_0(R_0)] \qquad (71)$$

$$E[B_0(R_0)] = \frac{1}{Q_0} \sum_{y=R_0+1}^{R_0+Q_0} E_{D_0(L_0)}[(D_0(L_0) - y)^+] = \frac{1}{Q_0} \sum_{y=R_0+1}^{R_0+Q_0} \cdot \sum_{u=y}^{\infty} (u - y) g_0(u)$$

(72)

Step 2.3.1 Estimate new induced backorder costs for VR_1 and optimize $R_{J=1}$. The approach is to treat VR_1 as a virtual central warehouse with lead time $\overline{L}_{j=1}$ satisfying direct customer demand and demand from VR_2 with different shortage costs and applying the iterative procedure in (Andersson et al 1998).

Step 2.3.2 Optimize R_1 for the backorder cost p_1, which is a function of β_1. First determine the induced backorder cost for VR_1 according to (73). Note that index 1 refers to VR_1 and index 2 refers to VR_2.

$$p_1 = \beta_2\left(\frac{\mu_2^{tot}}{\mu_1^{tot}}\right) + \widehat{p}_1\left(\frac{\mu_1}{\mu_1^{tot}}\right) \qquad (73)$$

Next is to determine the total lead time demand at VR_1, $D_1^{tot} = \sum_{j=1}^{2} D_j$, in units of Q. Remember that Q=1 is the largest common divisor of all order quantities. This can be done by exact convolution of the compound Poisson processes or by approximating the total lead-time demand by a negative binomial distribution with a mean of $\mu_1^{tot}\overline{L}_1$ and a standard deviation of $\sigma_1^{tot}\sqrt{\overline{L}_1}$ according to (74). The lead time demand distribution for VR_1 is denoted $g_1(u)$.

Let $P(D_1^{tot}(\overline{L}_1) = u) = g_1(u) \quad for\, u = 0, 1, 2...$ (74)

The cost function for VR_1 to minimize (75).

$$C_1(R_1) = h_1 QE[IL_1^+(R_1)] + p_1 QE[B_1(R_1)] \quad (75)$$

$$\text{where } E[B_1(R_1)] = E[IL_1^-(R_1)]$$

Simplifying the cost expression in (75), we get (76).

$$C_1(R_1) = \frac{(h_1+p_1)\cdot Q}{Q_1}\left\{\sum_{y=R_1+1}^{R_1+Q_i}\sum_{u=0}^{y}(y-u)g_1(u)\right\} - p_1 Q(R_1 + \frac{Q_1+1}{2} - \mu_1^{tot}\overline{L}_1)$$

(76)

$$R_1^* = max\left\{R_1 : C_1(R_1) - C_1(R_1 - 1) \leq 0\right\} \quad (77)$$

It is easy to show that, as $p_1 > 0$, $C_1(R_1) - C_1(R_1 - 1)$ is increasing in R_1, which implies that C_1 is convex. An optimal R_1^* that solves the problem and satisfies the optimality condition can therefore be found through a simple search according to (77).

If $R_1^* = R_1$, the reorder point corresponds to the previous set of induced backorder costs, and no further updating of the induced backorder costs and R_1 is needed.

Otherwise if $R_1^* \neq R_1 \Rightarrow$ set $R_1 = R_1^*$ and proceed to **step 2.4.1**

Step 2.4.1 Estimate the associated expected lead-time, $\overline{L}_2(R_1)$ according to the METRIC inspired approach according to (78).

$$\overline{L}_2(R_1) = \frac{1}{\mu_1^{tot}} \cdot E[B_1(R_1)] \quad (78)$$

Where $E[B_1(R_1)]$ is calculated according to (79).

$$E[B_1(R_1)] = \frac{1}{Q_1}\sum_{y=R_1+1}^{R_1+Q_1} E_{D_1^{tot}(\bar{L}_1)}[(D_1^{tot}(\bar{L}_1) - y)^+] = \frac{1}{Q_1}\sum_{y=R_1+1}^{R_1+Q_1} \cdot \sum_{u=y}^{\infty}(u - y)g_1(u)$$
(79)

Step 2.5.1 Determine a new estimate for β_2 based on R_1 and $\bar{L}_2(R_1)$. In principle, the same overall approach is used as in Berling et al. (2023) based on the iterative procedure in Andersson et al. (1998). That is, we assume demand is normally distributed with correct mean and variance, and estimate β_2 as the derivative of this retailer's expected holding and backorder costs with respect to its lead time, $\bar{L}_2(R_1)$, assuming R_2 is chosen optimally. It is shown that for normally distributed demand, an iterative procedure can be used to find an optimal β value for the approximation model (Berling and Marklund, 2006). For other demand distributions like the compound Poisson, the method is a robust approximation. An optimal R_2 is determined for which (80) is minimized. Here, $\phi(x)$ is the CDF of the normal distribution.

$$\beta_2(\bar{L}_2) = (h_2 + \hat{p}_2)\frac{\sigma_2^2}{\mu_2 Q_2}[\phi(\frac{R_2^* + Q_2 - \mu_2\bar{L}_2}{\sigma_2\sqrt{\bar{L}_2}}) - \phi(\frac{R_2^* - \mu_2\bar{L}_2}{\sigma_2\sqrt{\bar{L}_2}})] \qquad (80)$$

Once an optimal R_2 has been found, this is used to estimate a new induced backorder cost for VR_2 according to (81).

$$\beta_2^{new} = \frac{dC_2^A(R_2^*|\bar{L}_2)}{d\bar{L}_2}\cdot\frac{1}{\mu_2} = (h_2 + \hat{p}_2)\cdot\frac{\sigma_2^2}{2Q_2\cdot\mu_2}\left[\phi(\frac{R_2^* - \mu_2\cdot\bar{L}_2 + Q_2}{\sigma_2\cdot\sqrt{\bar{L}_2}}) - \phi(\frac{R_2^* - \mu_2\cdot\bar{L}_2}{\sigma_2\cdot\sqrt{\bar{L}_2}})\right]$$
(81)

If $\beta_2^{new} = \beta_2^{old} \Rightarrow$ stop updating β_2 and proceed to **step 2.6.1**

If $\beta_2^{new} \neq \beta_2^{old} \Rightarrow \beta_2 = \beta_2^{new}$ and proceed to **step 2.3.2**

step 2.6.1 Once the induced backorder cost, $\beta_{2,}$ for VR_2 stabilizes, use the new induced backorder cost for VR_1, VR_2, and the regular retailers to calculate the weighted induced backorder cost at the central warehouse. Given the new induced backorder cost, calculate a new reorder point for the central warehouse according to **step 2.1.3**.

Andersson et al. (1998) showed that for constant $p_i \geq h_i$ this iterative procedure is guaranteed to converge. The key property, which assures convergence of the procedure is the concavity of $C_i^A(R_i^*|\overline{L}_i)$ with respect to $\overline{L}_i$. A necessary but not very restricting condition for this property to hold is that the unit costs at retailer i must satisfy; $p_i \geq h_i$. A complete analysis of the relationship between the stationary solutions and the optimal solution to the problem in question, together with a thorough investigation of the behavior of the coordination procedure, can be found in (Andersson et al., 1998).

For the special cases when there are only two channels active, i.e. regular retailers and one virtual retailer, the iterative procedure presented above can be used with slight modification. By simply setting the demand for the VR_2 equal to zero. Another approach requiring less computation is to follow the steps presented in (Berling et al., 2023) for one virtual retailer. This approach is conceptually similar and uses the results in (Andersson et al., 1998) to find convergence.

5.3.3 Step 3. Determine the reorder points for regular retailers

Given the reorder point at the central warehouse, the reorder point for the regular retailers, can be determined. Note that for a given R_0 the reorder points, $R_{1,2,3\ldots N}$ can be optimized independently for each retailer. (Berling et al., 2023). This is done by solving the fill rate constrained singe-echelon problem (82). The objective is to find the lowest cost while fulfilling the target fill rate, γ_i^*. Recall that index i refers to regular retailers.

$$min\, C_i(R_i) \;=\; h_i E[(IL_i)^+] \;=\; \frac{h_i}{Q_i} \sum_{y=R_i+1}^{R_i+Q_i} E_{D_i(\bar{L}_i)}[(y \;-\; D_i(\bar{L}_i))^+] \quad (82)$$

$$s.t.\ \gamma_i(\bar{L}_i(R_0),\ R_i) \geq \gamma_i^*$$

where $\bar{L}_i$ is determined according to the METRIC inspired approach in (Berling et al., 2023), according to (83), where l_i is the transportation time from the central warehouse to retailer i, μ_0 is the demand at the central warehouse during the lead time, L_0, and $E[B_0(R_0^*)]$ is the expected number of backorders.

$$\bar{L}_i \;=\; l_i + \frac{L_0}{\mu_0} E[B_0(R_0^*)] \qquad (83)$$

Under the assumption of compound distributed demand, the fill rate, γ_i, is computed according to (84), where $f_i(d)$ is the probability of a customer ordering d units at retailer i.

$$\gamma(R_i,\ R_0) \;=\; \frac{\sum_{d=1}^{\infty} \sum_{j=1}^{\infty} min(j,d) \cdot f_i(d) \cdot P(IL_i = j \mid R_0, R_i)}{\sum_{d=1}^{\infty} d \cdot f_i(d)} \qquad (84)$$

The inventory level probability distribution $P(IL_i = j \mid R_i)$ is determined according to (85), where the distribution for the lead-time demand $D_i(\bar{L}_i)$ can be computed by exact convolution of the compound poisson distribution or by approximating the demand according to negative binomial distribution with the correct mean, μ_i, and standard deviation, σ_i.

$$P(IL_i = j \mid R_i) \;=\; \frac{1}{Q_i} \sum_{k=max(R_i+1,\, j)}^{R_i+Q_i} P(D_i(\bar{L}_i) \;=\; k \;-\; j)\ for\, j \;\leq R_i + Q_i$$

$$0\ otherwise \qquad (85)$$

5.3.4 Step 4. Determine reservation levels for virtual retailers

The first section describes in detail how the reservation level for VR_1 is determined, followed by a section describing how the reservation level for VR_2 is set.

5.3.4.1 Determine the reservation levels for VR_1

The inventory level for the combined stock at the central warehouse and VR_1 is described in (86). Note that the inventory level of the combined stock, $IL_{0,1}$ is equal to $IL_0 + S_1$ when the inventory level at the general stock is positive, i.e. $IL_0 > 0$. When $IL_0 \leq 0$ the combined stock on hand $IL^{+}_{0,1}$ is equal to IL^{+}_{1} and all stock is reserved for the direct upstream demand, i.e. demand entering the virtual channels. The probability distribution of the combined stock on hand, $IL_{0,1} \geq 0$, may then be determined according to (87).

$$IL^{tot}_{0,1} = \begin{cases} IL_0 + S_i & for\ IL_0 > 0 \\ IL_1 & for\ IL_0 \leq 0 \end{cases} \tag{86}$$

$$P(IL^{tot}_{0,1} = j) = \begin{cases} P\left(IL_0 = j - S_1\right) & if\ \ j > S_1 \\ \left(1 - P\left(IL_0 > 0\right)\right) \cdot P\left(IL_1 = j\right) & if\ \ j \leq S_1 \end{cases} \tag{87}$$

Once the probability distribution of the inventory level for the combined stock is determined, equation (88) can be used to determine fill rate, given S_1 and R_0.

$$\gamma(S_1,\ R_0) = \frac{\sum_{d=1}^{d_{max}} \sum_{j=1}^{R_0+S_1+Q_0} min(j,d) \cdot f_1(d) \cdot P(IL^{tot}_{0,1} = j \mid R_0, S_0)}{\sum_{d=1}^{d_{max}} d \cdot f_1(d)} \tag{88}$$

Where $f_1(d)$ is the probability distribution of customer order sizes at VR_1 excluding demand occurring at VR_2, and is determined according to (89), where O_1 is the size of customer order at VR_1.

$$f_1(d) = P(O_1 = d),\ d = 1, 2, \ldots d_{max} \quad (89)$$

To determine the base stock S_1 for a given R_0, the process involves incrementally raising this value from zero until the fill rate surpasses the specified target fill rate. Alternatively, if ready rate, RR, is used instead of fill rate, the calculation can be performed by assessing the probability function associated with the combined stock at the central warehouse and VR_1 being greater than zero, according to (90).

$$RR(R_0, S_1) = P(IL_{0,1}^{tot} > 0 \mid R_0, S_1) = \sum_{j=1}^{R_0+S_1+Q_0} P(IL_{0,1}^{tot} = j \mid R_0, S_1) \quad (90)$$

In the preceding sections, the determination of the inventory level distributions at the general stock and virtual retailers are explained.

5.3.4.1.1 Determining the inventory distribution for the general stock

The probability distribution of the central warehouse inventory can be computed according to (91).

$$P(IL_0 = j) = \frac{1}{Q_0} \sum_{k=max(R_0+1,m)}^{R_0+Q_0} P(D_0(L_0) = k - j \quad (91)$$

Here $D_0(L_0)$ denotes the subbatch demand (in units of Q) at the central warehouse during the lead-time L_0, and is computed according to (59) or (60) depending on distribution type. Recall that since the model assumes a base-stock policy at the virtual retailers, the system subbatch, Q, will always be equal to 1.

5.3.4.1.2 Determining the inventory level distribution for the virtual retailer

As previously mentioned, the inventory level distribution for the virtual retailer is inherently difficult to determine due to stochasticity in the delay caused by stock-outs of the general stock.

To arrive at an efficient approximation, note that delays only occur when the general stock is depleted, i.e. when $IL_0 \leq 0$. Let $\widehat{L}$ denote the expected lead-time for units ordered by the artificial retailer that experience a delivery delay. The probability of stock outs and the general warehouse is computed according to (92).

$$P(IL_0 \leq 0) = 1 - P(IL_0 > 0) = 1 - \sum_{k=1}^{R_0+Q_0} P(L_0 = k) \quad (92)$$

Focusing only on these situations and assuming that the lead-time is constant and equal to its mean, the inventory level for the virtual retailer can be computed according to (93).

$$IL_1 = S - D_1^{tot}(\widehat{L}) \quad (93)$$

where $D_1^{tot}(\widehat{L})$ denotes the stochastic lead-time demand at VR_1. In this case, the demand encompasses direct demand from customers supplied by VR_1, and also indirect demand from the proceeding virtual retailer, i.e. VR_2. $D_1^{tot}(\widehat{L})$ is computed according to (94).

$$D_1^{tot}(\widehat{L}) = \sum_{i=1}^{2} D_i(\widehat{L}) \quad (94)$$

The pmf of the lead-time demand can for the purpose of determining IL_1 be obtained by exact convolution or by a negative binomial approximation. If negative binomial approximation is used, the total lead-time demand is approximated with mean $\mu_1^{tot}\widehat{L}_1$ and a standard deviation of $\sigma_1^{tot}\sqrt{\widehat{L}_1}$.

The probability mass function (pmf) for the stock on hand at the virtual retailer is then determined according to (95).

$$P(IL_1 = j) \ = P(D_1^{tot}(\hat{L}_1) = S_1 - j) \qquad (95)$$

To estimate $\hat{L}_1$, it is assumed that the lead-time L_1 follows a two point distribution such that it is zero when there is stock on hand at the general stock, and $\hat{L}_1$ when there is a delay. The probability that there is a delay is denoted by α. Furthermore, it is assumed that the mean lead time $E[L_1]$ is equal to the expected delay per unit delivered from the central warehouse stock, $\overline{L}$, and is determined using Little's law, $\overline{L} \ = E[IL_0^-]/\mu_0$. Given these assumptions, $\hat{L}_1$ can then be determined according to (96).

$$\hat{L}_1 = \frac{\overline{L}_1}{\alpha} \qquad (96)$$

Note that in the scenarios under consideration, $\overline{L}_1$ may not represent the precise average delay encountered by units distributed to a specific retailer. Nevertheless, it typically serves as a robust approximation, as shown in (Berling et al., 2023).

The probability α may be estimated in different ways. (Berling et al., 2023) propose to set $1 - \alpha$ equal to the ready rate for the general stock at the central warehouse, RR_0, which by definition is the probability that $IL_0 > 0$. Note that this corresponds to the proportion of time that at least some part of an upstream demand order is satisfied

without a delay. This give the following estimation of $\hat{L}_1$ according to (97).

$$\hat{L}_1 = \frac{\overline{L}_1}{1 - P(IL_0 > 0)} \qquad (97)$$

5.3.4.2 Determine the reservation level for VR_2

The methodology for establishing the base stock policy for the second virtual retailer follows a similar process as that of VR_1, with slight modification.

When considering VR_2, which is replenished from VR_1, we can treat the combined stock of the central warehouse and VR_1 as a virtual central warehouse, with the inventory level $IL_{0,1}^{tot}$. We can then determine the combined stock level of VR_2, $IL_{0,2}^{tot}$, in the same way as for VR_1. $IL_{0,2}^{tot}$ is determined according to (98), and the inventory level distribution determined according to (99).

$$IL_{0,2}^{tot} = \begin{cases} IL_{0,1}^{tot} + S_2 & for\ IL_{0,1}^{tot} > 0 \\ IL_{0,2} & for\ IL_{0,1}^{tot} \leq 0 \end{cases} \tag{98}$$

$$P\left(IL_{0,2}^{tot} = j\right) = \begin{cases} P\left(IL_{0,1}^{tot} = j - S_2\right) & if\ \ j > S_2 \\ \left(1 - P\left(IL_{0,1}^{tot} > 0\right)\right) \cdot P\left(IL_2 = j\right) & if\ \ j \leq S_2 \end{cases} \tag{99}$$

When $IL_{0,2}^{tot}$ is determined, the smallest S_2 which fulfills the target fill rate according (100) is deemed optimal.

$$\gamma(S_2,\ R_0) = \frac{\sum_{d=1}^{d_{max}} \sum_{j=1}^{(\sum_{k=1}^{2} s_k)+R_0+Q_0} min(j,d)\cdot f_2(d)\cdot P(IL_{0,2}^{tot}=j|\ R_0,S_2)}{\sum_{d=1}^{d_{max}} d\cdot f_2(d)} \tag{100}$$

Note that when the lead time is approximated to VR_2, $\overline{L}$ is approximated with the total demand experienced at VR_1, and the expected number of backorders for VR_1, see (101). $\widehat{L}_2$ is the expected lead-time for units ordered by VR_2 that experience a delivery delay (i.e., a lead-time >0) because of stockouts. $\widehat{L}_2$ can then be approximated according to (102).

$$\overline{L}_2 = \frac{1}{\mu_1^{tot}} E[(IL_{0,1}^{tot})^{-}] \qquad (101)$$

$$\widehat{L}_2 = \frac{\overline{L}_2}{1 - P(IL_{0,1} > 0)} \qquad (102)$$

Chapter 6. Numerical study

The objective of the numerical study is aligned with the purpose of the master thesis stated in section 1.4.1 and revolves around examining the impact of applying the EM-model instead of the SE-mode currently used by Volvo. The study focus on Volvo's distribution system in South Africa.

The numerical study comprises two main sections. The first part examines situations when there is no service differentiation between the virtual retailers, i.e. the demand from VR_1 and VR_2 adhere to the same target service level. The second part explores the model's performance under three service differentiated channels, i.e. VR_1 and VR_2 have different target service levels. The latter phase serves as an exploration, considering potential benefits for Volvo in the future.

Volvo's distribution system requires a multi-echelon inventory control system capable of managing dealers under Volvo´s direct control and those outside their direct oversight. This setup is reflected in the model, with two distinct distribution channels from the CW: one for controllable retailers and another for non-controllable retailers integrated within the central warehouse stock.

Given the operational context, service differentiation across non-controllable retailers is not a primary concern, the necessity for an additional channel isn't deemed essential for this thesis project. However, Volvo is interested in exploring further service differentiation across multiple channels. Hence, the latter part of the numerical study aims to explore various scenarios where virtual retailers VR_1 and VR_2, serve demand from independent retailers with different fill rate requirements. The demand data for the system will be varied in order to test the model under more dynamic circumstances. For detailed information about the test data, see Appendix D.

The numerical study consists of two different discrete event simulation models in ExtendSim 10 for each item. The first simulation, acting as a base-line scenario, is henceforth denoted "SE-model". Input for the

SE-model consists of data including reorder point and order quantities obtained from Volvo's system. The second simulation represents the suggested EM-model. In this simulation, output from the first SE simulation as well as output data from the analytical model was used as input data. In the following section these two input sources and the ExtendSim 10 model are described in more detail.

6.1 Data collection and analysis

The initial step of the data collection and analysis was an initial clean up of the data. Mentioned in the delimitation in section 1.5 this master thesis will focus on the optimization between the Regional Distribution Center (RDC) in Johannesburg and dealers replenishing from this warehouse for a smaller set of articles in the GTO division. The replenishment to Johannesburg from the Central Distribution Center (CDC) in Gent is therefore from a modeling perspective seen as a replenishment from an outside supplier in the numerical study.

In collaboration with SML, a set of items were chosen and investigated in the numerical study. These items are characterized by varying demand dynamics, ranging from slow to fast movers exhibiting highly variable customer order sizes. A condition for the studied items is that a compound Poisson demand distribution is applicable. This facilitates a correct comparison between the analytical model and the extendsim simulation. Within this segment, different price classes were explored to encompass a broad spectrum of the items under investigation, ensuring analysis and representation of the current market dynamics. The choice of low and lumpy items is supported by two main factors. Firstly, such products pose significant control challenges compared to stable items with high demand with relatively low standard deviation. Secondly, simulating normally distributed demand, representing stable items, is problematic due to the discrete nature of demand against continuous distribution, rendering the solution difficult to validate in Extendsim. In Appendix E, the data for the different items is presented. As mentioned earlier, the choice of items should reflect the market dynamics, hence the selection of items ranging from lower demand to higher demand (from 0,76 units/day to 15,97 units/day). The numerical study encompasses a total of 15 items.

In Table 3, the input data for the analytical model and the simulation model is stated. This data was collected from DSP for dealers with LPA agreement and from an internal SQL database for dealers without LPA on the 22th of Mars 2024. The DSP contains data of customer orders, while the SQL contains data from the RDC. The collected data in DSP was aggregated on a monthly basis and represented data points for the previous year at the dealers. The corresponding data for the same period in SQL was collected at the RDC. Since the discrete event simulation in Extendsim is based on days, much of the data had to be converted to this format, see column "Data source" in Table 3 for conversion. Words in *italic* in Table 3 indicate that data is collected from DSP or SQL from a field with a corresponding name.

Table 3: Data source corresponding to the input data.

Input data	**Data source**	**Used in**
Transportation time for retailer i, l_i Initial time unit: weeks	*Lead time* / 7	SE-model Analytical-model
Mean demand per day for retailer i, μ_i Initial time unit: months	Average($\sum_{i=-12}^{-1} (Sales\ period\ i)/(4.3 \cdot 7)$)	SE-model Analytical-model
Standard deviation per day for retailer i, σ_i Initial time unit: months	Standard deviation($\sum_{i=-12}^{-1} (Sales\ period\ i)/$ $(\sqrt{4.3 \cdot 7})$)	SE-model Analytical-model
Order Quantity	EOQ	SE-model Analytical-model
Reorder point RDC	*Safety stock* +(*Forecast* * *Planning lead time* /(4.3 ・7))	SE-model
Reorder point retailers	(*Safety stock_new* + *Forecast* / 4,3) /*Lead time_Weeks*	SE-model

While retailers have the option to order any item from the RDC, not all items experience demand at every retailer. Consequently, for this study, we assume non-existent demand for items not registered in the inventory system at retailers where no sales have occurred. Retailers with existing demand and a variance-to-mean over one are set as active, this is compiled for the investigated items in Appendix E. The number of active retailers in the network for the different items is ranging between 10 and 14 retailers.

Returns registered in the inventory system are excluded from our investigation. We assert that efficient inventory control obviates the need for returns. In the case of returns in the inventory system, this is simply ignored. Similarly, emergency shipments, which entail shorter transportation times but higher costs, are disregarded based on the premise that a well-functioning inventory and distribution system negates their necessity. Notably, Volvo also currently lacks a fundamental logic for emergency shipments, rendering the decision parameters of our analytical and simulation models impractical in representing real-world scenarios.

6.1.1 Service differentiated channels

The implemented EM-model can handle three service differentiated channels, as illustrated in Figure 15.

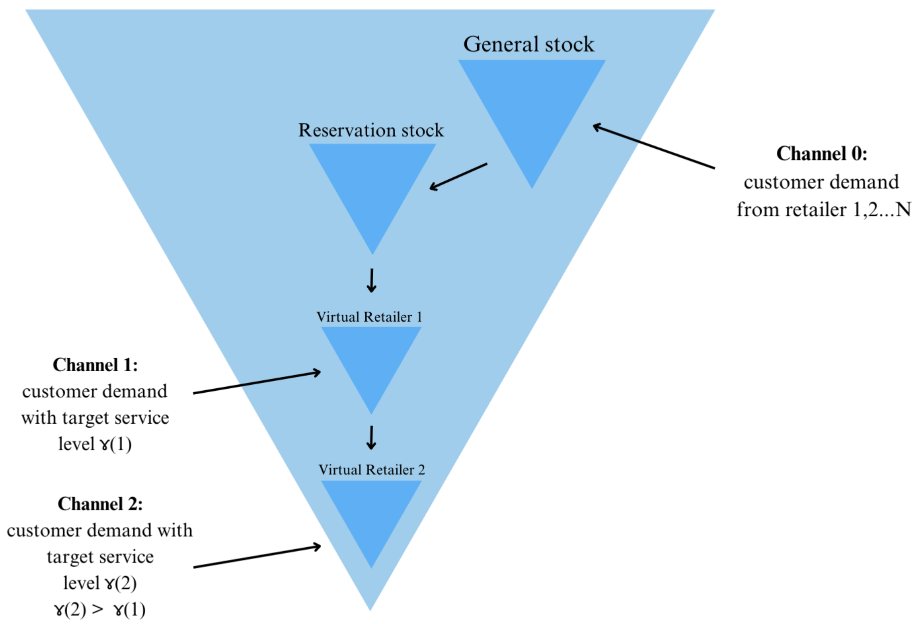

Figure 15: Service differentiated channels in the EM-model.

The customer demand originates from two distinct sources: the network of retailers denoted as i = 1,2,... N, all under the control of Volvo, and the capacitated retailers from channels virtual retailer 1 and 2, with increasing target service levels. Demand at the RDC coming from the regular retailers is denoted channel 0. The demand at the RDC originating from VR_1 and VR_2 is denoted channel 1 and 2 respectively. Within the context of Volvo, the notation outlined in Table 4 will be employed.

Table 4: Description and notation of the different demand channels.

	Description	**Notation**
Regular retailers (Channel 0)	These are the retailers owned by Volvo, whose inventory policy Volvo can control with an LPA-agreement.	Regular retailers
Virtual retailer 1 (Channel 1)	These are the retailers without an LPA-agreement, whose inventory policy Volvo can not control.	VR_1
Virtual retailer 2 (Channel 2)	These are retailers with an LPA agreement, but not owned by Volvo.	VR_2

In the initial phase of the numerical study, no distinction in service level between VR_1 and VR2 will be implemented. The demand will continue to be allocated between the two channels based on its origin. Nevertheless, with identical service level targets, the outcome should indicate that no reservation stock is required for VR_2. The decision to maintain separate demand channels stems from the realization that combining them would yield close to identical results, with the slightest deviation of the approximated demand distributions. Our intention is to investigate the model's performance in the subsequent phase, where service differentiation between the two channels is introduced.

The mean and variance of the demand at the virtual retailers is compiled as the sum of the ingoing dealers mean demand and variance.

6.1.2 Distribution fitting

The data collected from Volvo present challenges in determining demand distributions. Sales history is aggregated monthly, providing information on the number of picks and total units sold during each period, typically expressed in months. This poses difficulty in accurately determining the distribution of customer order sizes due to the limited granularity of the data. To address this challenge, instead of attempting to estimate the order size probability distribution for each retailer individually, a logarithmic compounding distribution is employed. This method is commonly used in practice when detailed information about order sizes is unavailable. However, the nature of the order distribution in the logarithmic model can result in a notable probability of very large order sizes, which may not accurately reflect real-world ordering behavior and consequently lead to inflated inventory levels required to meet target fill rates.

Nonetheless, for the purpose of this master thesis, which aims to compare inventory control systems, this assumption regarding demand distribution is not restrictive. This is because both systems operate under the same assumption regarding demand distribution. Therefore, while the use of the logarithmic compounding distribution may introduce some inaccuracies in representing real-world order sizes, it does not impact the comparative analysis of inventory control systems, as both are subject to the same assumption.

For dealers without an LPA agreement, Volvo lacks data beyond the orders they have placed at the RDC, essentially limited to batch records at specific time stamps. Given the scarcity of information regarding the order probability distribution, we had to resort to rudimentary assumptions. Following discussions with our supervisor, Prof. Johan Marklund, we interpreted the orders logged at the RDC as indicative of demand. While this interpretation may not precisely reflect reality, it's an acceptable approximation given the absence of detailed information about the demand at these retailers.

To fit the collected demand data to a logarithmic compounding distribution, the formulas outlined in Section 4.2.2 are employed. It is important to note that for the distribution to be valid, the variance-to-mean ratio must exceed one. However, this requirement does not pose a limitation for the items under investigation.

6.2 Analytical model

The analytical model was implemented in Python, the preferred programming language at Volvo. The EM-model was developed following the specifications outlined in Section 5.3.

It's important to highlight that, at this stage, the analytical model necessitates manual intervention. Data essential for computing inventory policies is supplied as inputs in an Excel spreadsheet, from which the Python-based analytical model retrieves and utilizes for optimization.

The negative binomial approximation for demand during the lead time for retailers, can pose limitations due to computational constraints. For high values of r, see (20), combined with relatively high demand, the computation of the probability mass function can be time consuming because values tend towards infinity. An alternative approach is to find a better demand distribution than the logarithmic distribution to model the customer order sizes. In cases where the problem took longer to solve, there were probabilities of very large customer order sizes, which is definitely not the real case for the item under investigation. In these scenarios, we opted to remove the item from the study since the analytical model in Python can not calculate the fill rate given the assumptions of the demand distribution.

6.2.2 Using the single echelon-model as a base-line comparison

As previously outlined, the SE-model serves as Volvo's existing single-echelon inventory control system, forming the basis for evaluating the analytical model. For a detailed understanding of the SE-model's operations, see Section 2.1.

To ensure that the study's results are both comparable and valid, it is imperative that the two models operate under uniform assumptions, thereby mitigating any biases within the systems. It is noteworthy that the optimized parameters derived from the SE-model undergo a different process than our analytical model, potentially resulting in service level disparities in the simulation compared to the fill rates in Volvos inventory control system. This discrepancy primarily stems from differences in assumptions regarding demand distribution and the SE-model's single-echelon optimization.

To overcome this, we adopt an approach where the simulated fill rates, derived from input parameters retrieved from Volvo's system, serve as the target fill rates for the analytical model. This ensures comparability across results and allows us to evaluate the EM-model's ability to achieve fill rates with less inventory in the system. The input parameters were obtained from databases within Volvo, provided by our supervisor at Volvo, regarding both the RDC and the retailers for the selected articles. See Figure 16 for an illustration of the process.

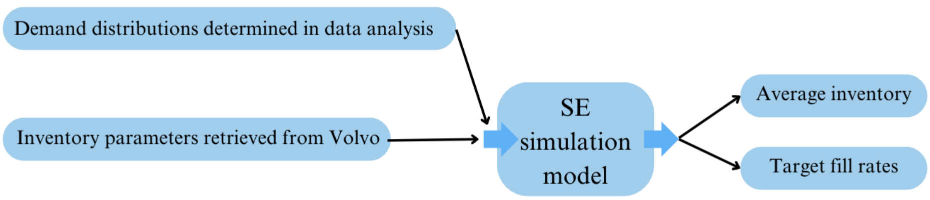

Figure 16: The process of obtaining average inventory and target fill rates in the SE simulation model.

6.3 Discrete event simulation

An essential part of the numerical study involves conducting discrete event simulations, using the specified inventory parameters to assess the fill rate and associated inventory levels within the system. This section provides a more in-depth description of the ExtendSim 10 model developed for this purpose.

Employing a discrete event simulation model enables evaluation of the EM-model, under comparable conditions with the SE-model. The input parameters concerning demand, order quantities, transportation times, holding costs, and backorder costs remain consistent across both models. The optimal reorder points derived from the analytical model are used as inputs, yielding average fill rates and inventory levels over the multiple runs.

6.3.1 Model setup

The discrete event simulation model used in this master's thesis was a modified version of a research model previously developed in ExtendSim 10. This model was developed by the division of Production Management at the Faculty of Engineering, Lund University. In the modified version, additional retailers were added to the original version. A conceptual setup of the modified model is illustrated in Figure 17.

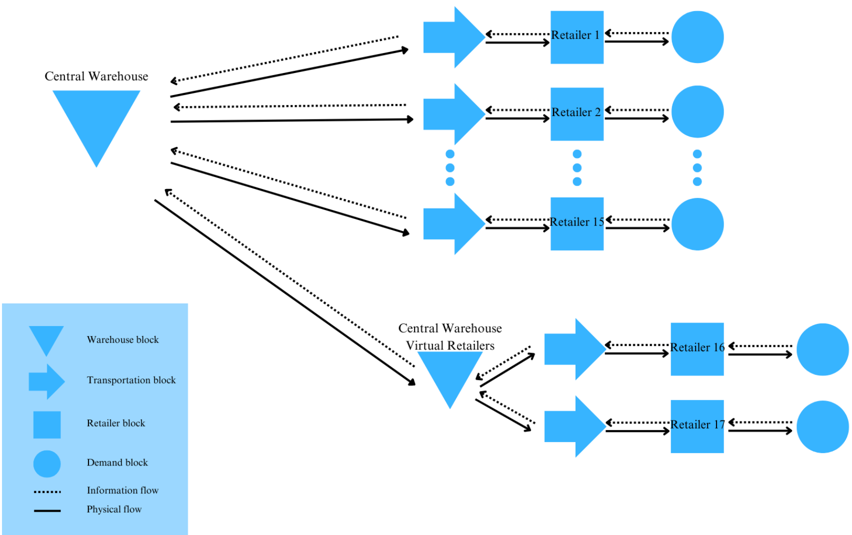

Figure 17: A conceptual overview of the discrete event simulation in ExtendSim 10.

The model replicates the one warehouse multiple retailers (OWMR) structure observed in the market replenishing from the RDC in Johannesburg. Consisting of two warehouse blocks, this setup allows us to mirror the combined stock policies in the EM-model. The virtual retailers, labeled as 16 and 17 in the model, replenish from virtual warehouses, which in turn are restocked from the central warehouse with a transportation time of zero. Conversely, the regular retailers replenish from the central warehouse.

Demand enters the model via a demand block, assuming a compound Poisson distribution. By assigning a parameter specifying the originating retailer, the system ensures that the demand is directed to the correct dealer. In cases where the system only experiences demand from VR_2 and not from VR_1 , VR_2 is modelated as VR_1. As a result of this, the VR_2 channel is only active when the system experiences demand in all three distribution channels.

The transfer of information, depicted by the dotted line in Figure 17, is transmitted throughout the model without any time delay. Inventory parameters utilized as input data in the model are managed within the ExtendSim database. Additionally, this database facilitates exporting results for further analysis in Excel.

The chosen items undergo two simulations each utilizing reorder points from the SE-model and the EM-model optimization, respectively. Each simulation consists of 50 runs over a period of 4100 time units for each run. At the onset of each run, all installations possess a stock-on-hand equivalent to the maximum stock of $R_i + Q_i$ units, with a warm-up period set to 250 time units. The warm-up period ensures that the fact that the system starts with maximum stock does not have an impact on the result and data is only collected after the warm-up period is reached.

6.3.2 Model validation and verification

The study culminated in an analytical model which already had a validated and verified ExtendSim model. The simulation model was developed by researchers at the Department of Industrial Management and Logistics at Lund University. The models support a structure of one warehouse and multiple retailers (OWMR), where the warehouse supplies the downstream retailers.

Adjustments were made to tailor the model to the South African market, including the addition of more retailers and updates to the demand distributions. To validate the model, test cases were conducted by significantly reducing system variability. This approach aimed to ensure that the simulated results closely matched those of the analytical model. By minimizing randomness, the model's behavior became more predictable, facilitating validation. Additionally, the mean values of the demand in the simulation were compared with those of the analytical model under equivalent conditions to ensure consistency. The verification of the simulation model in ExtendSim found a discrepancy in the logarithmic generating demand block, compared to theoretical values. Therefore, the probability of different customer order sizes had to be manually inserted, thereafter resulting in a fully functioning simulation model.

Chapter 7. Result and analysis

The analysis of the results will be divided into two parts. The first part, section 7.1, is the service differentiation across two channels, and the second part, section 7.2, is the result of service differentiation across three channels. The analysis will focus on how well the model achieves the target fill rates and examine the expected inventory levels in the system. Here the base-line, as discussed in section 6.2.2, is the outcome from the simulation based on parameters from the currently used SE-model. This is then compared with the simulation outcome using the parameters from the analytical model.

For the second part, no valid comparison can be made between the current practice and the EM-model. Therefore, this part will focus mainly on how well the model achieves the desired fill rates for the different channels.

7.1 Service differentiation across two channels

In this section, we will focus on presenting and discussing the results concerning service differentiation across two specific channels. To clarify, this pertains to both regular retailers and additional channel(s) that share the same target fill rate. Consequently, both VR_1 and VR_2 may be operational, yet adhere to identical service level targets.

7.1.1 Reorder points

This section examines the proposed decision parameter, the reorder points, across various installations. The analysis aims to show the potential systematic changes in decision-making that may arise with the implementation of the EM-model in the Volvo distribution process.

The reorder point at the RDC is presented in Figure 18, and the reorder points for the dealers are presented in Figure 19. In both Figures, the current reorder points for dealers and the optimized reorder points using the EM-model are present.

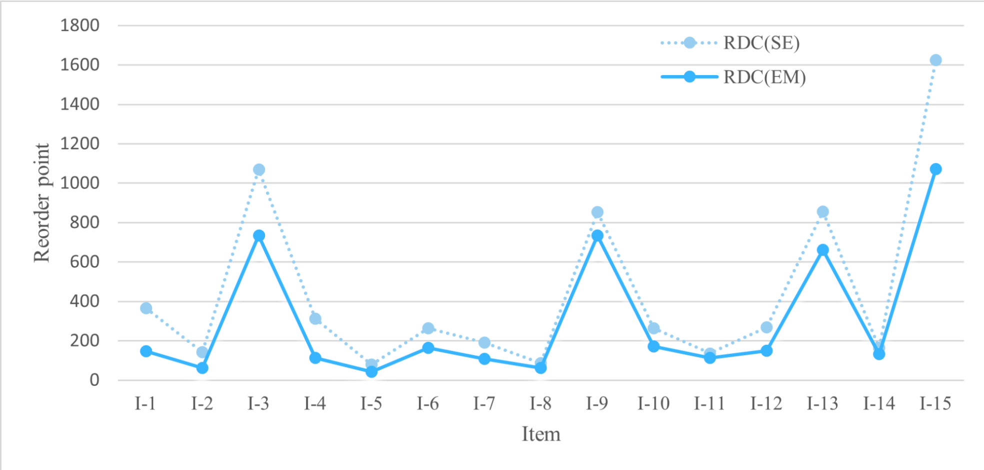

Figure 18: Reorder point at the RDC for the different items.

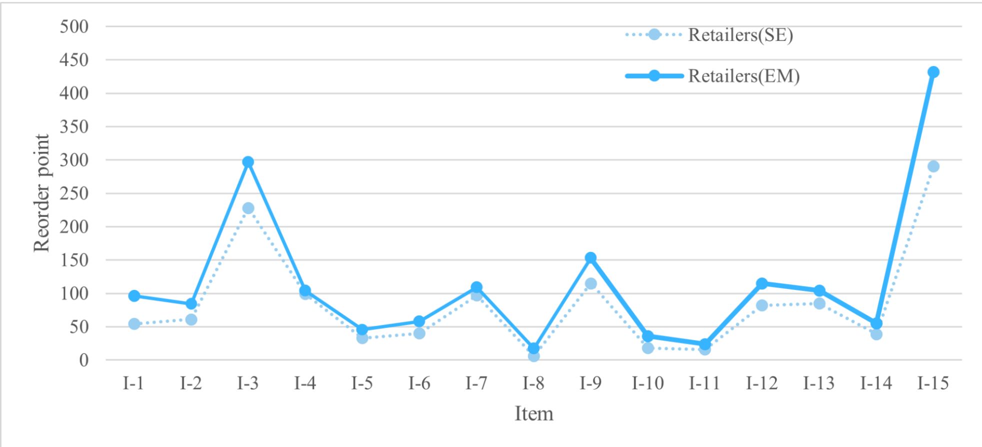

Figure 19: Reorder point at the regular retailers for the different items.

Figure 18 shows that the reorder point at the RDC in Johannesburg was decreased in all cases. In total the reorder point was decreased by 33 percent. Moreover Figure 19 shows that the aggregated reorder point at the retailers increases in all cases. In total the aggregate reorder point for retailers was increased by 37 percent for the different items. This result is expected since as previously mentioned multi-echelon models such as the EM-model tend to push stock downstream in the system resulting in higher reorder points at retailers.

Figure 20 illustrates the reorder points for virtual retailers, all sharing the same target service level in this phase of the study. Remember that the reorder points for virtual retailers are the reorder points for the reservation stock dedicated to serve the virtual channels, i.e. this is added to the RDC stock.

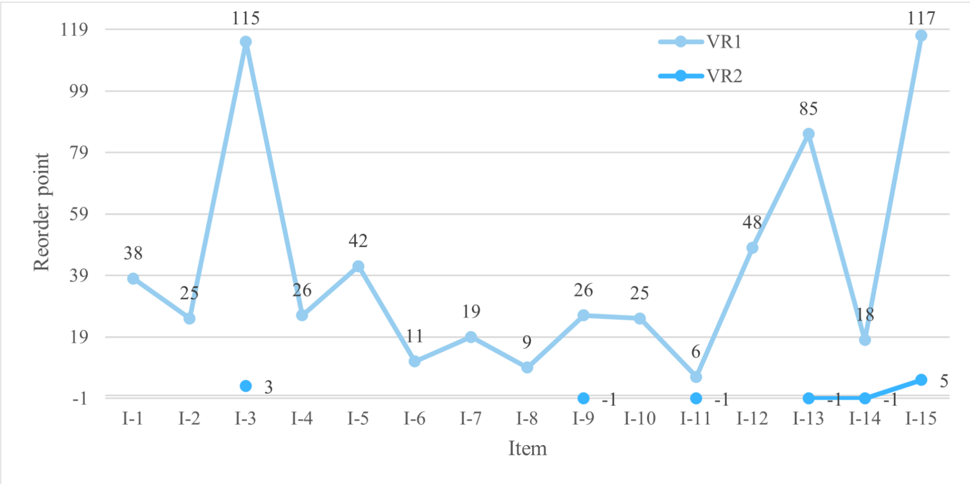

Figure 20: Reorder point for virtual retailers.

In the majority of cases, when only VR_1 has a reservation level, VR_2 remains inactive, experiencing no demand for the item under investigation. Conversely, when both virtual retailers are active, most results indicate a reservation level (R+1) of zero for VR_2, signifying that no additional reservation stock is needed to fulfill VR_2's demand. This aligns with expectations, as VR_1's reservation level already encompasses the demand to be supplied downstream to VR_2, and since both channels share identical service levels, no further reservation is deemed necessary.

However, in certain instances, VR_2 necessitates additional reservation stock to achieve desired fill rates. This observation underscores that despite both virtual channels aiming for the same service level, supplementary reservation stock for VR_2 becomes important for fulfilling fill rate requirements. This finding is interesting, suggesting that even with uniform service targets across channels, varying demand sources may mandate reservation levels to attain desired fill rates. From modeling perspective, demand adhering to the same target service level can be bunched together in the virtual channels. The case when combining the demand from VR_1 and VR_2 into one channel becomes interesting, and is illustrated in Table 5.

Table 5: Results from treating VR_1 *and* VR_2 *as either separate or combined for item 6.*

Test	**Mean demand (μ) and standard deviation (σ) per time unit**	**Reservation levels**	**Ready rate**	**Fill rate diff (pp.)**
VR_1 and VR_2 treated as separate channels	VR_1: $\mu = 1{,}24$; $\sigma = 3{,}52$ VR_2: $\mu = 3{,}29$; $\sigma = 7{,}29$	$S_1 = 116$ $S_2 = 4$	$VR_1 = 96{,}7\%$ $VR_2 = 97{,}1\%$	$VR_1 = -4{,}92$ $VR_2 = -0{,}84$
VR_1 and VR_2 treated as combined channels	VR_{com}: $\mu = 4{,}53$; $\sigma = 8{,}10$	$S_{com} = 116$	$VR_{com} = 96{,}7\%$	$VR_{com} = -4{,}99$

As seen from Table 5, when VR_1 and VR_2 are analyzed individually, the deviations from the target are -4.92 and -0.84, respectively. However, when considered as a combined demand source, the deviation is -4.99. Given that the reservation level remains identical for both the combined scenario and VR_1 in isolation, it is not surprising that the deviations are nearly identical in both cases. The differentiating factor lies in the reservation level for VR_2, which consequently leads to additional stock and thus a higher fill rate for VR_2. The model still does not fully achieve the target fill rate for VR_2, but the deviation is smaller.

The result indicates that further segmentation between channels adhering to the same service level can result in higher fill rates for the proceeding virtual retailer. However, this is a special case of when the demand is weighted differently in the two virtual channels and the analytical model suggests further stock allocation to VR_2. However, the majority of items investigated in this study suggests a reservation level for VR_2 set to zero under the circumstances of identical service level targets. This indicates that bunching them together or treating them as separate will in most cases achieve the same fill rate. The deviation in fill rate can rather be explained by the lead time approximation, as further elaborated in section 7.1.2.1.

7.1.2 Fill rates

Firstly the average fill rate difference from the base-line across all retailers is presented in Figure 21. The result indicates that the difference for most of the items is small with a deviation less than one percent points (pp). However for item 8, 10 and 11 the difference is relatively high with a higher average fill rate deviation of 12,3, 6,3 and 6 percent respectively. For other items, for example 1 and 5 the average fill rate difference is low, under 0,2 percent.

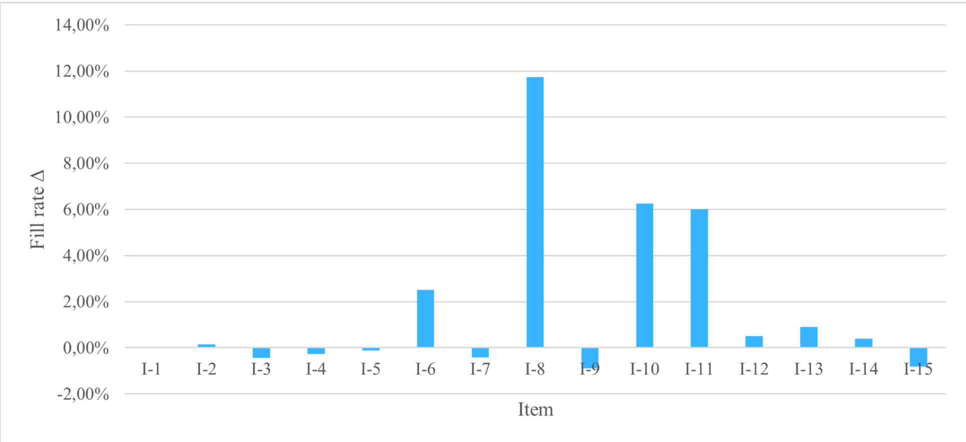

Figure 21: Average fill rate difference for the different items for all retailers.

Figure 22 presents a box chart depicting the fill rate difference. The mean, indicated by a cross, represents the average fill rate difference shown in Figure 21. The colored box in Figure 22 represents the interquartile range, encompassing 50% of the observations. The T-shaped lines, or whiskers, denote the minimum and maximum values within the dataset. Any dots outside the whiskers are outliers, indicating values that significantly deviate from the rest of the data.

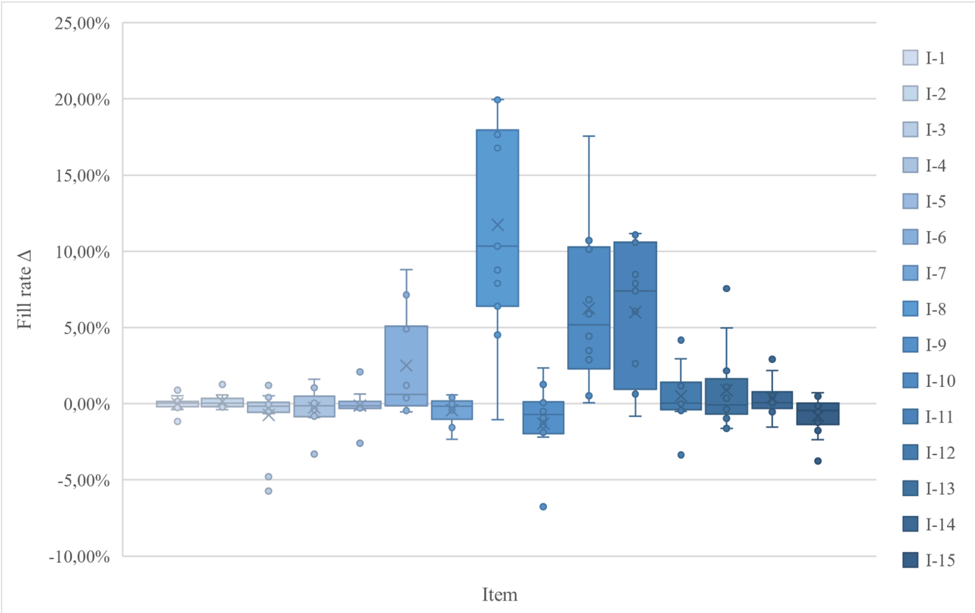

Figure 22: A box chart over the fill rates for the different items.

Overall, the fill rate difference for each item is close to zero with some instances where the fill rates are exceeded considerably. The overshoot can be explained by the discretization of the reorder points further elaborated in 7.1.2.1. According to the results, the analytical model performs well when achieving the target fill rates.

7.1.2.1 Fill rates for selected items

In this section the fill rate difference of item 1 and 8 is presented in more detail, corresponding to the items with the lowest and highest deviation in fill rate respectively.

Figure 23 illustrates the deviation in fill rates for item 1 across various retailers. Among them, retailers 3, 12, and the virtual retailer exhibit deviations exceeding 0.5 percent, with disparities of 0.53%, 0.88%, and -1.16%, respectively. Despite the presence of both positive and negative variations, along with relatively minor differences, the average deviation for this item, as depicted in Figure 21, remains low. This outcome suggests that while the overall discrepancy is low, certain retailers may experience higher deviations in their fill rates. The latter can not be certain by only looking at the mean fill rate achieved over multiple simulation runs. For retailer 12 and VR_1, with the highest deviation from target fill rate, the standard deviation of the mean fill rate is 2,5 and 1,1 percentage points respectively. This indicates that we can not say by statistical significance that the expected fill rate deviates from the target, considering a 95% confidence interval.

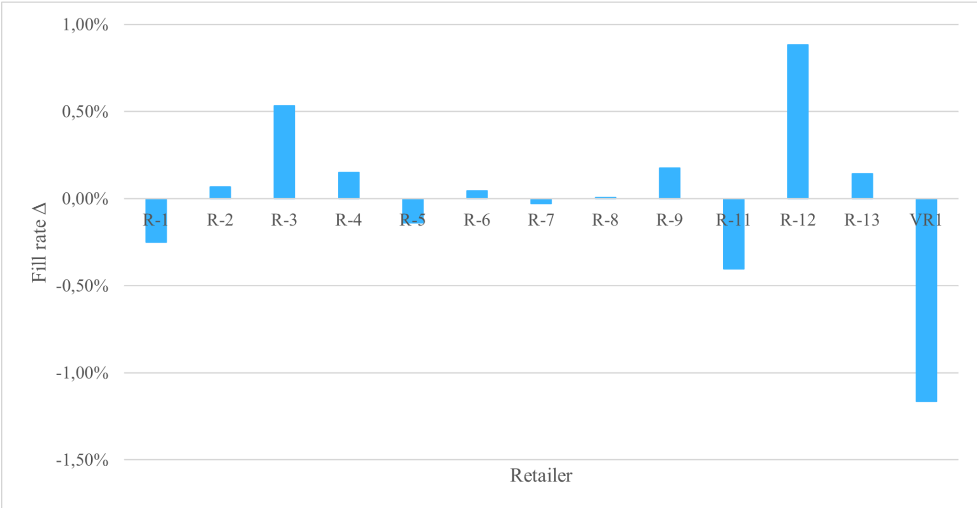

Figure 23: Fill rate difference for item 1 at the different channels, $\sigma_{R\text{-}12}$=2,5 pp and σ_{VR1} = 1,1 pp.

Figure 24 shows that the fill rate difference for item 8 is consistently high, overshooting the fill rate by 4,5 to 20 percentage points, at all active retailers except the first virtual retailer. This explains why the average fill rate difference for this item is relatively high in Figure 21.

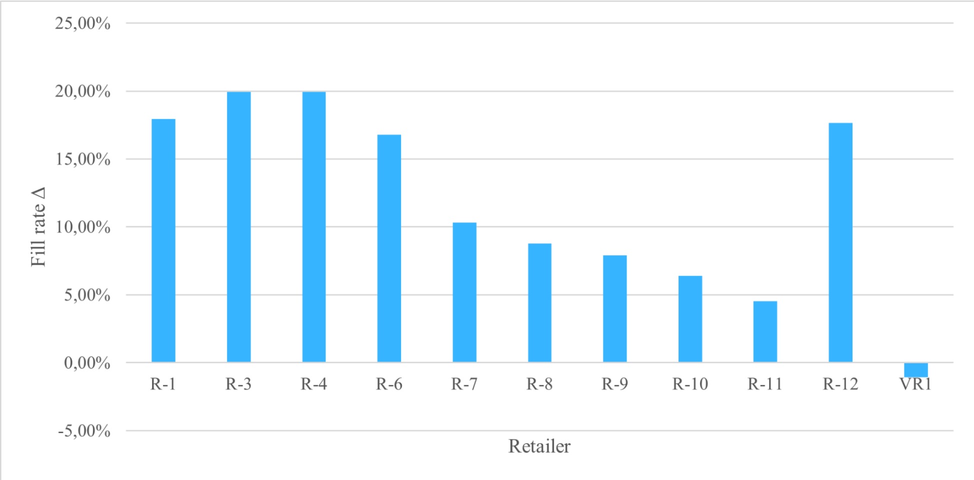

Figure 24: Fill rate deviation for item 8 at the different channels.

This is a result of the discretization of the reorder points. For items experiencing low demand, have low ordering quantities, and require a relatively low target fill rate, the decision of the reorder point can influence the fill rate significantly, as illustrated in Figure 25. Figure 25 showcases the obtained fill rate for different reorder points for retailer 1, here the average demand is 0,006 units/day with a standard deviation of 0,105 units/day. Consequently, in order to attain the target fill rate, the actual fill rate must sometimes surpass expectations considerably.

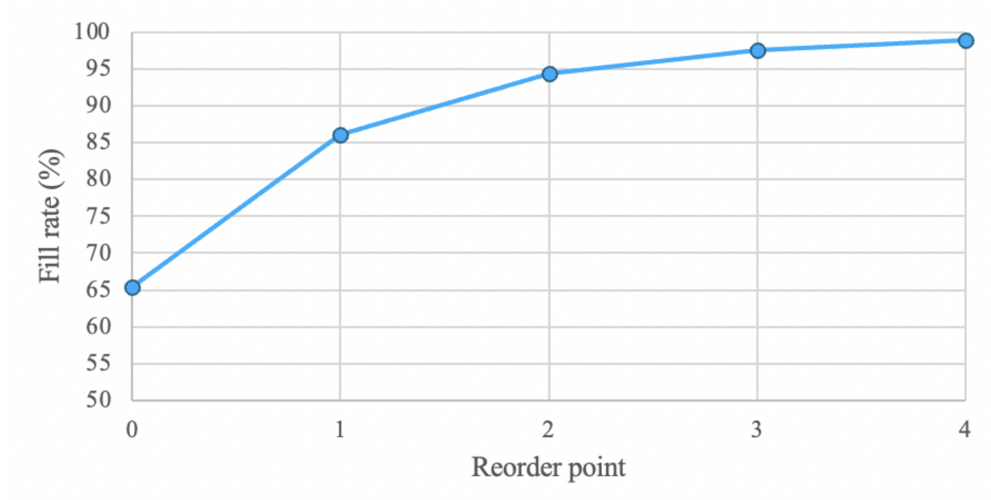

Figure 25: Decision of reorder points and the impact on the fill rate for retailer 1.

Showcased in Figure 26 this is not only true for item 8 for retailer 1. In Figure 26 the reorder point is reduced by one at all retailers where the fill rate difference is positive. Instead of exceeding the target fill rate a majority of the retailers falls below the target fill rate. This indicates that the model successfully sets the reorder point to obtain the target fill rate.

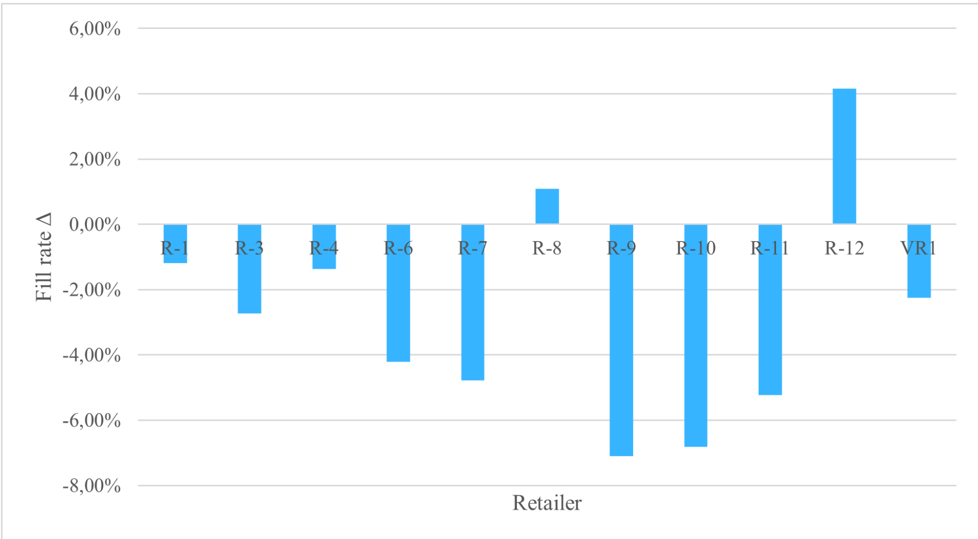

Figure 26: Fill rate deviation for item 8 when reorder point is reduced by one at all installations.

7.1.2.2 Fill rates for virtual retailers

In this section the fill rates for the virtual retailers is examined in more detail. Figure 27 below presents a box chart over the fill rate difference for the virtual retailers over the items. The achieved fill rates for virtual retailers slightly come up short of the targets. However, the mean deviation for all items is less than two percentage points. In Figure 27, the mean is illustrated as a "x", the median as a solid line (slightly above the mean for VR_1 and VR_2), and the highest and lowest values illustrated by the T-shaped borders of the Figure.

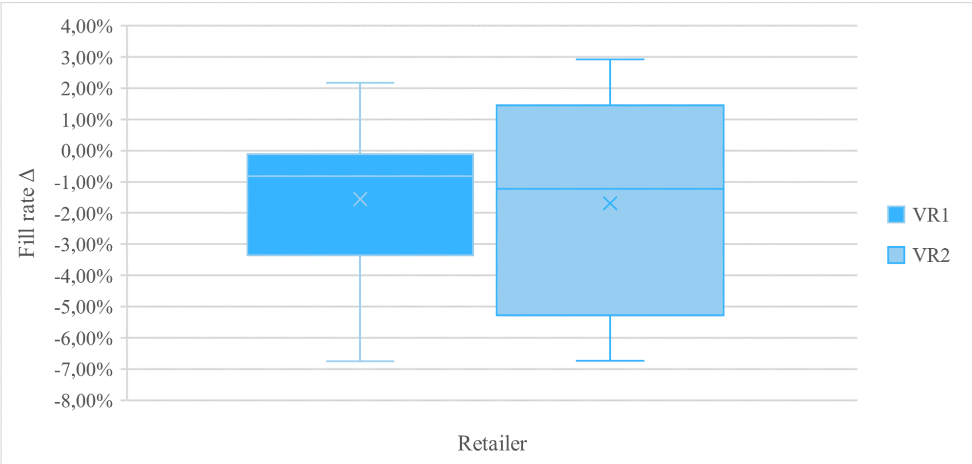

Figure 27: A box chart over the fill rate difference for the virtual retailers.

Delving deeper into the simulation outcomes, examining the confidence intervals is crucial for assessing the statistical significance of any fill rate deviations. If the confidence interval spans above and below the target range, we cannot confidently determine whether the model consistently meets or fails to meet the target fill rate.

Table 6 shows the mean fill rate difference and the standard deviation for the Virtual retailers across different items. Out of 15 simulated items, only two items (3 and 9) have a statistically significant deviation from the target fill rate, as their 95 percent confidence intervals are below the target service level. Deviations for all items are within a confidence interval of three standard deviations.

Table 6. Investigation of confidence interval for virtual retailers

Item	**Mean fill rate difference VR_1 (%)**	**Standard deviation VR_1 (pp)**	**Mean fill rate difference VR_2 (%)**	**Standard deviation VR_2 (pp)**
1	-1,16	1,1	-	-
2	-0,4	2,5	-	-
3	-5,72	2,0	-4,79	1,96
4	-0,81	1,1	-	-
5	-0,26	1,7	-	-
6	-0,44	0,9	-	-
7	-2,33	2,2	-	-
8	-1,05	2,3	-	-
9	-6,75	2,8	-6,74	2,8
10	0,06	2,1	-	-
11	0,63	2,0	0,96	1,9
12	-3,35	2,1	-	-
13	-0,12	0,8	-0,09	0,7
14	2,17	2,7	2,92	2,6
15	-3,75	3,2	-2,36	3,1

7.1.2.2.1 Lead time impact on fill rate

In this section, the lead-time to the RDC is investigated further to see if it has an impact on the achieved fill rate for the virtual retailers.

The test in Table 7 was conducted using a dataset encompassing diverse demand patterns, varying order quantities, and predetermined fill rates. In each test, the transportation time to various retailers varies, yet remains consistent across all trials. The target fill rates in the tests for VR1 and VR2 are 95% and 96% respectively. The creation of fictional demand patterns serves the purpose of encompassing a range of scenarios and evaluating fill rate deviations within the simulation. As elaborated in section 7.1.2.1, the reorder level significantly influences fill rates across different demand patterns, prompting the exploration of factors that may affect the model through the testing of diverse scenarios.

Table 7: Test data to investigate the lead time approximation.

Test case	**Lead time to RDC**	**Mean demand at retailers per time unit**	**Standard deviation at retailers per time unit**	**Order quantity retailers**	**Order quantity RDC**	**FR VR1 (%) , σ(pp)**	**FR VR2 (%) , σ(pp)**	**Difference from target FR (pp)**
1	84	0,4	1	1	68	93,1 2,1	95,0 1,9	**VR1:** -1,9 **VR2:** -1
	20	0,4	1	1	68	94,6 1,4	94,7 1,6	**VR1:** -0,4 **VR2:** -1,3
	10	0,4	1	1	68	94 1,4	96,3 1,4	**VR1:** -1 **VR2:** +0,3
2	84	0,6	1	3	45	90,2 2,4	93,4 2,1	**VR1:** -4,8 **VR2:** -2,6
	20	0,6	1	3	45	93,2 1,3	93,1 1,4	**VR1:** -1,8 **VR2:** -2,9
	10	0,6	1	3	45	92,2 1,2	95,7 1,2	**VR1:** -2,8 **VR2:** -0,3
3	84	1	2	5	140	91,5 2,2	92,8 2,2	**VR1:** -3,5 **VR2:** -3,2
	20	1	2	5	140	93,4 1,4	94,7 1,4	**VR1:** -1,6 **VR2:** -1,3
	10	1	2	5	140	93,6 1,4	94,9 1,4	**VR1:** -1,4 **VR2:** -1,1

Table 7 indicates that increased demand variability, stemming from prolonged lead times, leads to larger deviations from the target fill rates at VR_1 and VR_2. The increased variability in demand contributes to an increased frequency of stock-outs, consequently rendering lead times more erratic and resulting in higher deviation in the fill rates. The same tendency can be observed when increasing the demand variability for the same warehouse lead-time. Conversely, as seen in Table 7, the METRIC-inspired approximation serves as a robust estimate of the lead times when the transportation time to the RDC is relatively short, as illustrated by the lower fill rate deviation.

However, drawing any conclusions that the fill rate deviation is correlated to the lead-time approximation is riskful, and we reserve this section as one possible explanation for the relatively small deviations. It is worth highlighting that despite relying on METRIC-type approximations, the lead-time approximation effectively achieves target fill rates for the regular retailers. This can possibly be attributed to the impact of the relative error incurred when approximating stochastic lead times through the inclusion of stock-out delays. With pre-existing transportation time already factored in, this relative error becomes less pronounced. Conversely, for virtual retailers with a transportation time of zero, the relative error in lead time approximation has a higher impact on the fill rate.

7.1.3 Expected inventory levels

This section investigates the differences in expected inventory levels, to evaluate if cost benefits arise with the EM-model in comparison to the current SE-model used for inventory control at Volvo. In Figure 28, the total expected inventory is presented for each item. Each bar has a dotted area, representing the expected inventory at the controllable Volvo owned retailers. The rest of the bar represents the expected inventory at the RDC, including separate stock for virtual retailers. With reorder points set according to the analytical model, the average inventory in the system decreases by 24,62%. This outcome is unsurprising, as multi-echelon models typically drive stock downstream while concurrently reducing upstream inventory levels. The rationale behind this trend lies in the optimization of inventory allocation in the system. The expected inventory includes the general stock used to serve all channels, and the reservation stock used to serve direct demand stemming from independent dealers (virtual retailers).

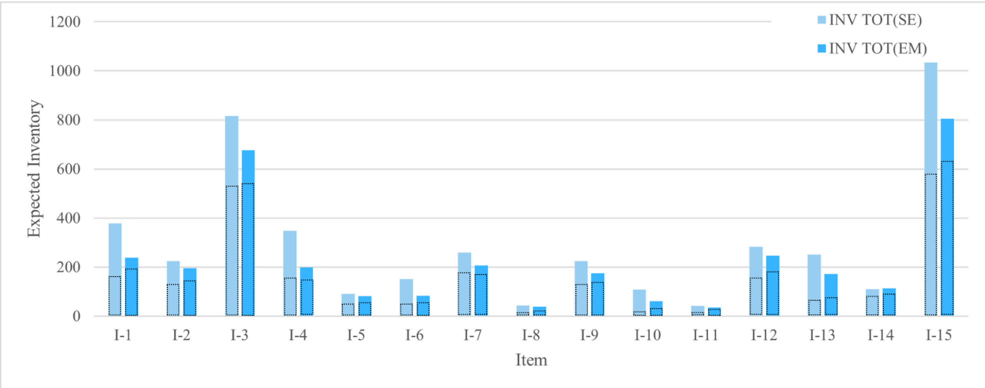

Figure 28: Expected inventory in the system for the different items.

7.2 Service differentiation across three channels

This section presents the findings of the exploration study on implementing service differentiation across three channels. While this practice is not currently applied within GTO at Volvo, examining its potential benefits offers valuable insights. The model performs well for regular retailers, as indicated in Section 7.1.2, with very low deviation from the target fill rates. Therefore, the result and analysis will solely focus on how well the model achieves the target fill rates for the virtual retailers.

The result of the study consists of five different test cases for two pairs of target fill rates. The target fill rate is illustrated by a horizontal line in each Figure. In the first test case a scenario where the variance-to-mean ratio of the demand is close to one is examined. The second case on the other hand examines a scenario where this ratio is higher. Case three examines the model's performance when most of the demand enters the system through VR_1. In case four a similar case is investigated where this demand arrives to the system though VR_2. The last case investigates a case where a majority of the demand arrives to the system through the regular retailers. In Appendix D the input data for the different test cases is stated. All test cases were conducted with a lead time to the RDC of 84 days and a fixed order quantity of 68 units at the RDC. Since the study is an exploration of the potential benefits of service differentiation in the Volvo distribution network, using the current lead time mimics the real system better. In Figure 29-32, μ represents the total mean of the demand, σ represents the standard deviation of the total demand, and S represents the reservation stock.

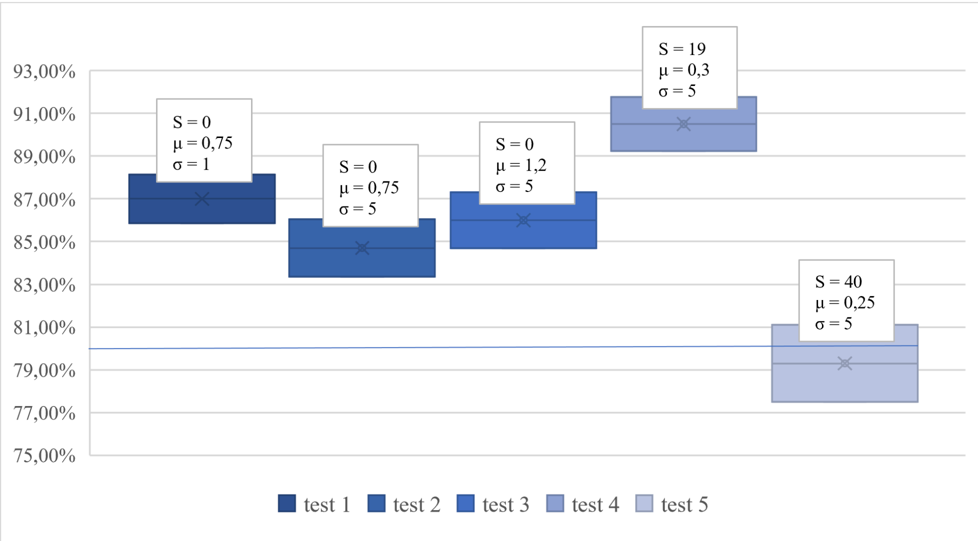

Figure 29: Fill rate mean for VR_1 ± *3σ with a target fill rate of 80%*

Figure 29 shows that the actual fill rate is above the target for the majority of tests. For test 1 through 3 the reservation level is set to 0, which means no reservation stock is necessary for VR_1. The resulting fill rate is the consequence of only relying on the general stock for supplying this channel, which in these cases have a fill rate exceeding the target fill rate. For test four the reservation level is set to 19 resulting in an overshoot of the target fill rate of roughly 10 percent. The demand in channel VR_1 is low, only 15 percent of the total demand and 60 percent of the total demand stemming from VR_2. When the majority of the demand stems from the regular retailers, the fill rate (±*3σ*) encapsulates the target.

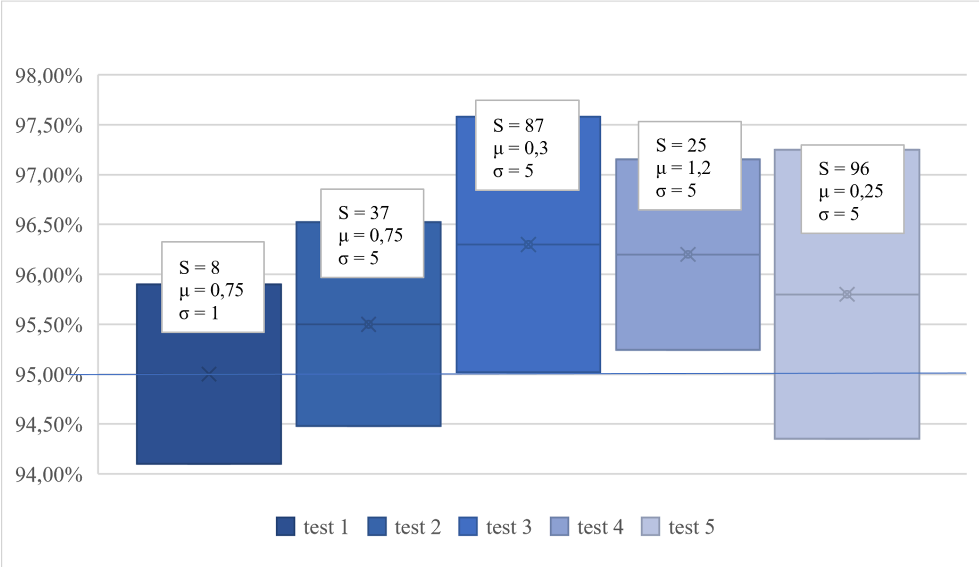

Figure 30: Fill rate mean for $\mathrm{VR}_2 \pm 3\sigma$ *with a target fill rate of 95%*

In Figure 30, the attained fill rate for VR_2 is depicted. The model successfully reaches the target fill rates across all test cases, with a slight overshoot for test four. For test case four, the relative overshoot for VR_2 is less than for VR_1 showcased in Figure 29.

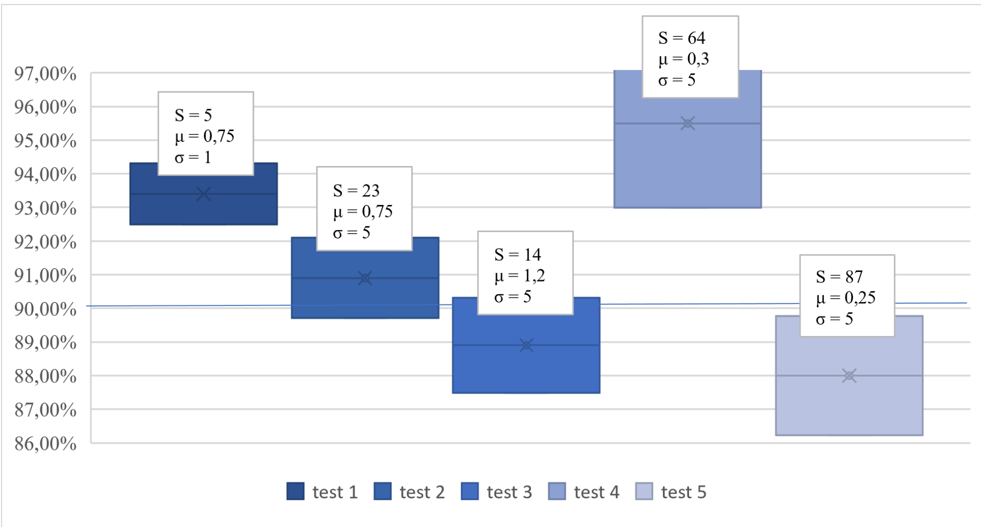

Figure 31: Fill rate mean for $\mathrm{VR}_1 \pm 3\sigma$ *with a target fill rate of 90%*

Figure 31 shows the fill rates for VR_1 for the different cases. As seen, in four out of five tests, the fill rate is within the interval of the expected fill rate. In the test where the fill rate falls short of the target, the deviation is -2 percentage points. Interestingly, in test four when the majority of the demand stems from VR_2 and a minority from VR_1, there is once again an overshoot of the target fill rate. In this case, an overshoot of roughly 6 percent. The model's performance when the majority of demand stems from VR_2 is further investigated in section 7.2.1.

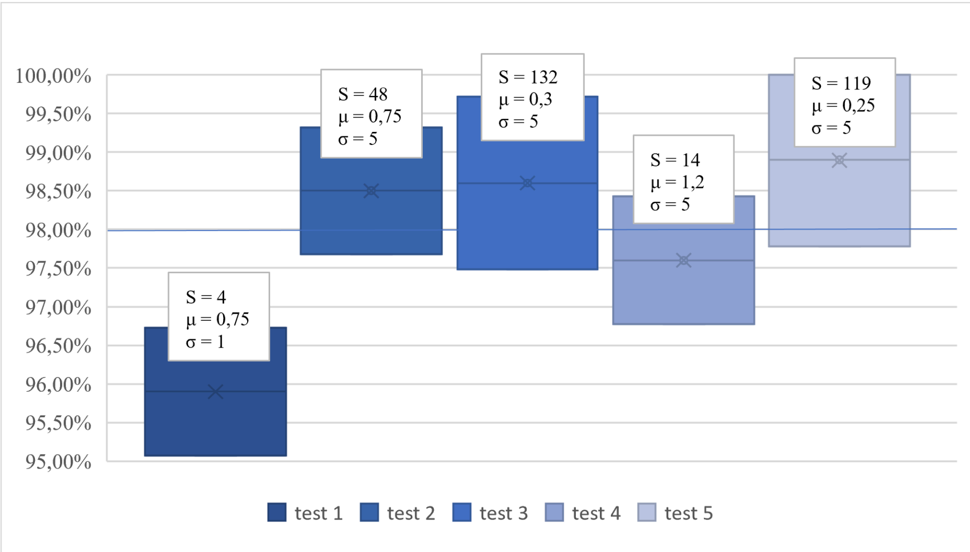

Figure 32:Fill rate mean for $VR_2 \pm 3\sigma$ *with a target fill rate of 98%*

As shown in Figure 32, the model successfully achieves the target fill rate for four out of five tests, with an undershoot of 2 percentage points for test 1.

One observation is that the model tends to achieve the target fill rate for VR_2 with less deviation from target. A possible explanation is, as the model is set up, the succeeding virtual retailers have higher target service levels. As illustrated in Figure 25, the reorder points have a bigger impact on the fill rate at lower values. Conversely, at higher reorder points, often translating to higher service level targets, the impact is less, as seen by the derivative in Figure 25.

This behavior can partly explain why the deviation for VR_2 is less than for VR_1, for both sets of service level targets, as seen in both Figure 33 and Figure 34. Additionally, when the target fill rate increases for VR_1 the deviation becomes less, as seen in Figure 34 compared to Figure 33.

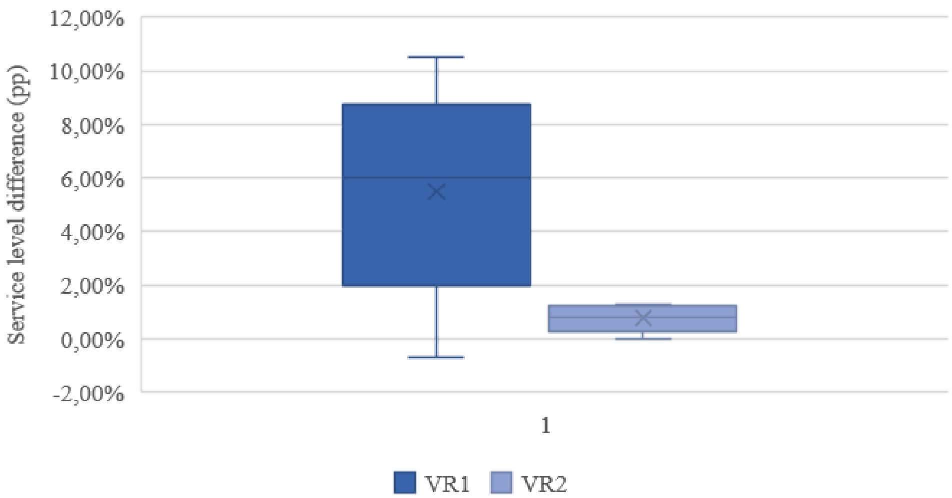

Figure 33: Service level difference for all tests when target fill rate is 80% and 95% for VR_1 and VR_2 respectively.

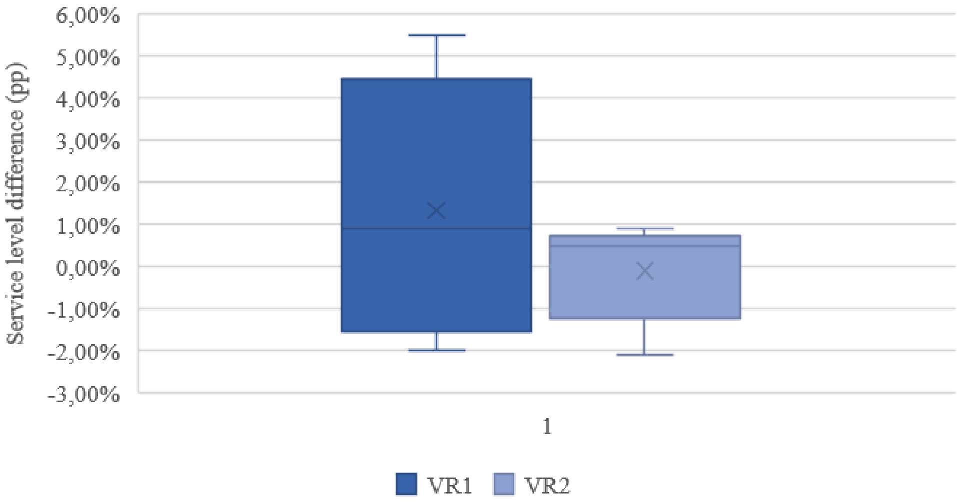

Figure 34: Service level difference for all tests when target fill rate is 90% and 98% for VR_1 and VR_2 respectively.

7.2.1 Performance of model with majority of demand in VR_2

In the initial test cases presented in section 7.2, it was observed that the average fill rate is above the target for VR_1 when most of the demand comes from VR_2. In this section more test cases are investigated to see if the trend remains. It is important to keep in mind that these tests have the purpose of examining the model's performance in an extreme situation. The test cases are presented in Table 8. In this section, the average fill rate difference for the regular retailers are presented as well. The variance-to-mean ratio in all test cases is intentionally set high, exceeding 80 in all test cases.

The majority of the demand is channeled through virtual retailers, representing direct demand at the central warehouse. Direct demand at the central warehouse diminishes the effects of a multi-echelon model because it bypasses the intermediary stages that characterize such systems. The purpose of a multi-echelon model is to distribute inventory across controllable retailers and the RDC. When the majority of demand is modeled as direct demand, the benefits of inventory allocation within the multi-echelon system are reduced. However, analyzing these cases may be valuable, as some items from Volvo, not investigated in the first section of the results and analysis, may exhibit these patterns.

Table 8: Data for test cases when the majority of demand stems from VR_2.

Test case	**Mean demand and standard deviation for regular retailers per time unit. (μ, σ)**	**Mean demand and standard deviation for virtual retailers per time unit. (μ, σ)**	**Demand in each channel (%)**	**Target FR**
4.1 (same as test 4)	0,0555; 1	VR1:0,3; 5 VR2:1,2; 5	Regular: 25% VR1: 15% VR2: 60%	VR1: 80% VR2: 95%
	0,0555; 1	VR1:0,3; 5 VR2:1,2; 5	Regular: 25% VR1: 15% VR2: 60%	VR1: 90% VR2: 98%
4.2	0,0555; 1	VR1: 0,5 ; 5 VR2: 1,5 ; 5	Regular: 20% VR1: 20% VR2: 60%	VR1: 80% VR2: 95%
	0,0555; 1	VR1: 0,5 ; 5 VR2: 1,5 ; 5	Regular: 20% VR1: 20% VR2: 60%	VR1: 90% VR2: 98%
4.3	0,0555; 1	VR1: 0,5 ; 5 VR2: 1 ; 5	Regular: 25% VR1: 25% VR2: 50%	VR1: 80% VR2: 95%
	0,0555; 1	VR1: 0,5 ; 5 VR2: 1 ; 5	Regular: 25% VR1: 25% VR2: 50%	VR1: 90% VR2: 98%

The resulting fill rate from the tests for VR_1 and VR_2, are presented through Figure 35 to Figure 38.

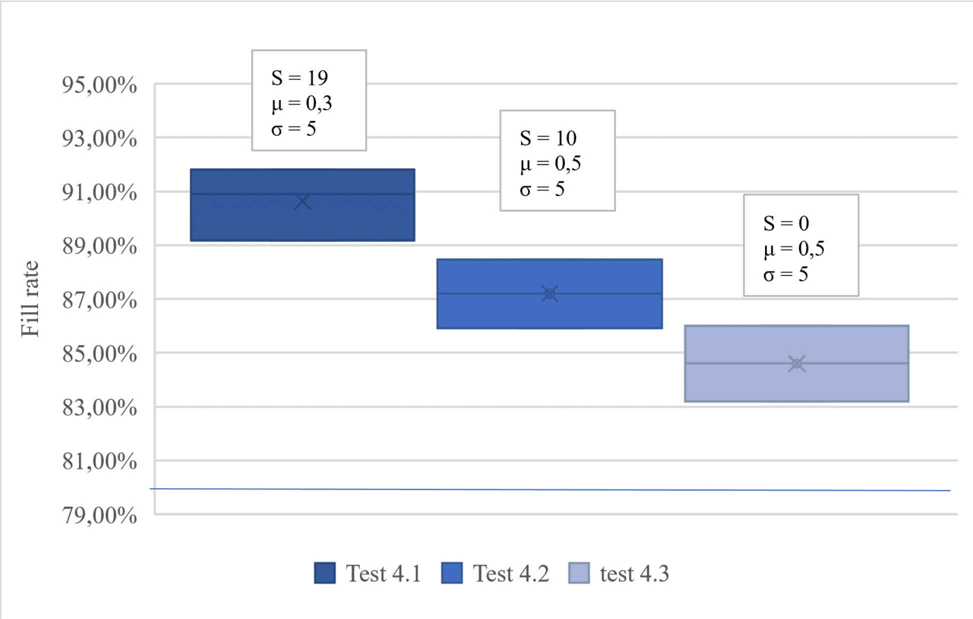

Figure 35: Fill rate mean for $VR_1 \pm 3\sigma$ *when target fill rate is 80%.*

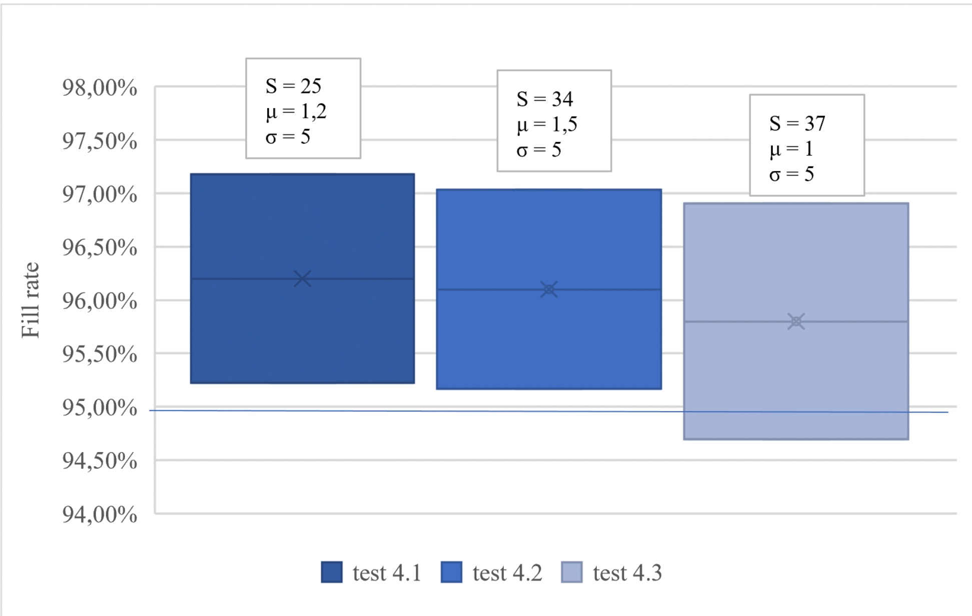

Figure 36: Fill rate mean for $VR_2 \pm 3\sigma$ *when target fill rate is 95%.*

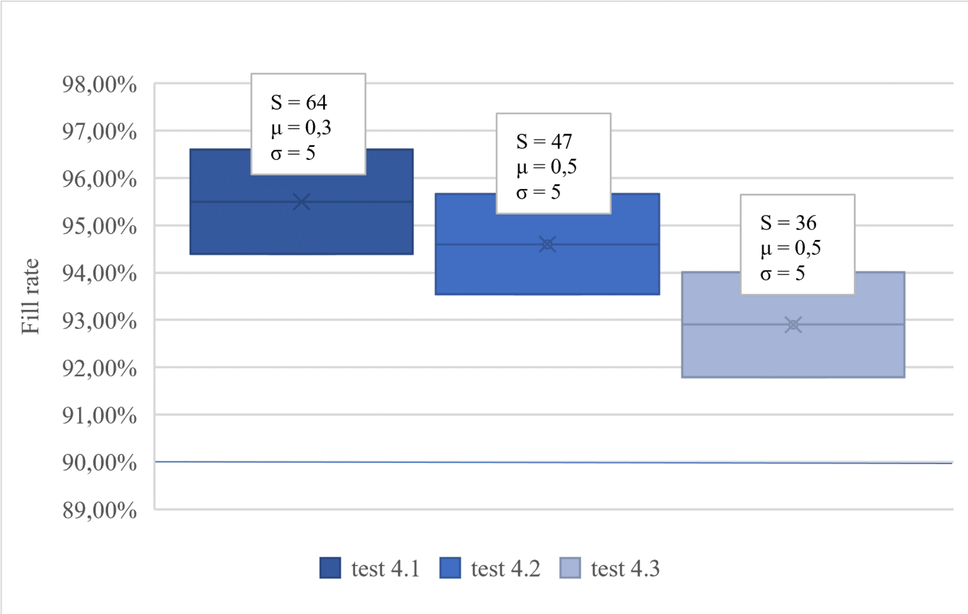

Figure 37: Fill rate mean for $VR_1 \pm 3\sigma$ *when target fill rate is 90%.*

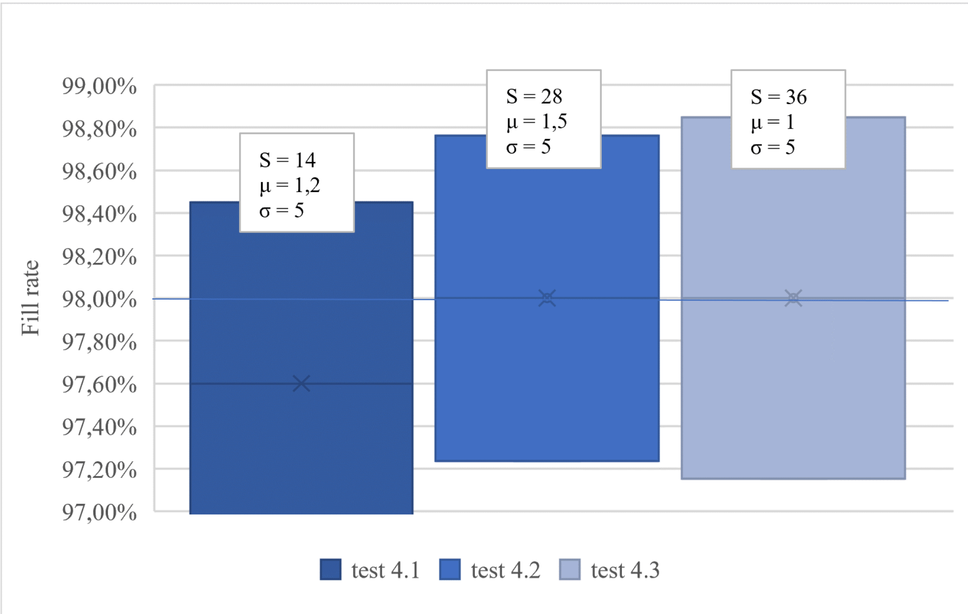

Figure 38: Fill rate mean for $VR_2 \pm 3\sigma$ *when target fill rate is 98%.*

As can be seen from Figure 35 to 38, the overshoot for VR_1 remains. However, the deviation from the target fill rate becomes less when the target fill rate increases and/or the variance-to-mean ratio decreases, see test 4.1 and 4.2. This result is expected as the variance-to-mean ratio decreases, the demand gets more predictable and stable. However, for test 4.3, the variance-to-mean ratio remains the same for VR_1 compared to test 4.2. The parameters changed are the mean and standard deviation of demand for VR_2. The result yields a lower fill rate deviation for VR_1 with zero reservation stock for this channel. The achieved fill rate is therefore a result of the general stock at the RDC. Given the reorder point and order quantity at the RDC, the lowest achievable fill rate for VR_1 in test 4.3 is therefore roughly 85%, corresponding to the fill rate at the RDC.

However, for these test series it is interesting to find the true optimal policy at the virtual retailers in order to achieve the target fill rates. This was conducted for test 4.1, with the highest deviation in fill rate for both VR_1 and VR_2. The optimal policy was found by decreasing the reservation level for VR_1 until the fill rate aligns with the target. Given the optimal policy for VR_1 the optimal reservation level for VR_2 was found in a similar manner.

There are many combinations of S_1 and S_2 which satisfies the target fill rates, and finding the true optimal policy in the simulation model requires modifying reorder points for both regular retailers and the RDC. This is deemed too time consuming. Therefore, Table 10 presents a few local optimal policies which satisfies the target fill rates. In Table 9, the result for test 4.1 with policies from the analytical model is presented in more detail.

Table 9: Detailed data from analytical model for test 4.1.

Target fill rate	**Inventory policy for VR_1 and VR_2**	**Resulting fill rate given analytical model policies**	**Fill rate RDC given analytical model policies**	**Expected inventory in the system**
$VR_1 = 80\%$ $VR_2 = 95\%$	$S_{VR1} = 19$ $S_{VR2} = 25$	$FR_{VR1} = 90{,}5\%$ $FR_{VR2} = 96{,}2\%$	$FR_{RDC} = 86{,}4\%$	E[IL] = 437 units

Table 10: Local optimal inventory policies for VR_1 and VR_2 for test 4.1.

Target fill rate	**Local optimal inventory policies for VR_1 and VR_2**	**Resulting fill rate given optimal policies**	**Fill rate RDC given optimal policies**	**Expected inventory in the system**
$VR_1 = 80\%$ $VR_2 = 95\%$	$S_{VR1} = 0$ $S_{VR2} = 33$	$FR_1 = 85{,}9\%$ $FR_2 = 96{,}0\%$	$FR_{RDC} = 86{,}4\%$	E[IL] = 428 units
$VR_1 = 80\%$ $VR_2 = 95\%$	$S_{VR1} = 10$ $S_{VR2} = 25$	$FR_{VR1} = 89{,}5\%$ $FR_{VR2} = 95{,}7\%$	$FR_{RDC} = 86{,}9\%$	E[IL] = 428 units

From Table 9 we can conclude that the best local optimal policy for VR_1 and VR_2 given the circumstances for test 4.1 is a reservation level of zero for VR_1, with additional 33 units reserved for VR_2. The expected inventory in the system is reduced by nine units compared to the analytical model. This is a relatively small improvement compared to the reductions presented in section 7.13 where the EM-model was compared to the SE-model.

The findings suggest that the achieved fill rate overshoots the target when a majority of the demand in the system comes from the VR_2 channel. The EM-models fundamental concept is to reduce the fill rate at the RDC in order to achieve the inventory reduction benefits. When the majority of the demand is modeled as direct demand, the fill rate at the RDC increases in order to maintain the desired fill rate for the virtual channels. This is a consequence of the induced backorder costs which are demand weighted, according to (51). With higher demand in the virtual channels, more stock is required at the general stock in order to decrease the total cost of the system, hence resulting in a higher fill rate. However, using a multi-echelon model when most demand enters the system directly at the RDC undermines its ability to control retailers. In such cases, a single-echelon model is likely sufficient and more computationally efficient. However, Volvo's SKU data for the investigated items shows that most demand comes from controllable retailers rather than direct demand from non-controllable retailers, justifying the use of a multi-echelon model.

Still, our conclusion is that the model can handle the extreme situation described above. From a practical standpoint, it is better to overshoot rather than undershoot the desired fill rate. The optimal policy derived via reducing the reservation level in the simulation yielded results with a reduction of roughly 2% of the total expected inventory compared to the analytical model.

Another noteworthy observation emerges when comparing test 3 in section 7.2 with test 4.1, particularly in the context of total demand originating from virtual channels being kept the same. The notable distinction lies in the primary source of demand: test 3 predominantly originates from VR_1, whereas test 4.1 sees the majority stemming from VR_2. It's worth noting that in test 3, where VR_1 is the primary source, deviations from the fill rate are considerably smaller. Conversely, test 4.1 highlights that larger deviations tend to occur when the majority of demand emanates from VR_2. In test 3, there is no reservation stock required for VR_1 in order to achieve the target fill rate indicating that the model finds an optimal policy for VR_1.

The test cases are too few to draw any well-grounded conclusions. However, an interesting observation is that when determining the reservation levels for VR_1, the demand from VR_2 is considered. From the model´s perspective, VR_1 supplies VR_2 and the dedicated stock for VR_1 is used by VR_2 as well. When there is a large portion of demand stemming from VR_2, a big portion of the reservation stock for VR_1 is solely used to supply demand from VR_2. The necessary reservation stock to serve direct demand from VR_1 is relatively small. With the added reservation stock for VR_1, it is not surprising that the fill rate exceeds the target considerably.

However, putting the result from the analytical model and the optimal policy found in the simulation model in context, the relative decrease in total inventory in the system is very small. For practical reasons, the discrepancy in results would not have a big impact on the total cost.

7.3 Validity of the results

The numerical study was conducted on a limited sample size, making it necessary to conduct more testing on a larger number of items to draw any definitive conclusions. This additional testing is essential to validate the expected stock reduction and meet target service levels. The sample size aimed to represent a broad spectrum of products by considering the mean demand experienced by the entire system, different coefficients of variation (i.e., the degree of demand fluctuation), and the distribution of demand across different channels. With the broader spectrum of items investigated, one can cautiously assume similar results for a wider sample size, thus ensuring that the conclusions remain reliable and trustworthy.

The extensive manual work required for setting up the simulations was a limiting factor. All data had to be manually extracted from databases, which involved determining the distribution of order sizes to be used as input in the simulation model. Therefore, the sample size that was possible to consider within the time frame of this project was limited. As the distribution system expanded to include more active retailers, the simulation duration increased significantly. Extendsim is an excellent tool for validating results, however our application suffers from scalability issues.

The use of a logarithmic distribution for order sizes does not accurately reflect actual customer order sizes. Due to the lack of precise information, assumptions had to be made. In some cases, the logarithmic distribution led to a relatively high probability of very large customer order sizes, which were unrealistic for the items under investigation. These probabilities of very large customer orders makes it very sensitive to reach the target service levels. If a customer orders a large batch of items, even with a very low probability, it significantly impacts the fill rate attained over multiple runs.

This can explain the large variability in average fill rates. If the correct distribution of demand were available, the simulation model and the analytical model would likely produce better results with less variability and differences between experienced averages and targets.

Each simulation run represents roughly 11 years, and each item was simulated over 50 runs to accurately estimate expected inventory levels and fill rates. The simulation results assume a steady state in the system, but in reality, demand patterns are likely to change over such an extended period. Therefore, it is important to note that fill rates in practice will likely fluctuate considerably over shorter time spans. Currently, Volvo does not monitor fill rates at dealers, but the results from the analytical model will provide valuable insights for developing near-optimal inventory policies based on current demand conditions.

Chapter 8. Conclusion

The problem formulation presented in section 1.4.1 revolved around investigating the impact of applying a recently developed service-differentiated multi-echelon omnichannel inventory control model at Volvo.

Based on the conducted literature review and a mapping of the current inventory control system at Volvo the authors suggest that the EM-model is the best currently available model for the company meeting the set up requirement of service differentiation. This is also supported by the numerical results from a study of Volvo´s distribution system in Southern Africa showcasing promising results.

On average the expected inventory in the system is decreased by 24,62% for the investigated items, which in practice would lead to considerable system cost savings. The reduction in expected inventory can be explained by the reorder point reduction at the RDC when the system is optimized according to the EM-model rather than the SE-model. This could be accomplished while still maintaining or closely achieving (within less than one percentage point difference) the average fill rate of the currently adopted inventory control model.

In the case of negative fill rate deviation, this can be partially attributed to a trend of negative deviations observed at the virtual retailers. In all cases where the average fill rate difference for the items is negative, the negative deviations at the virtual retailers is bigger. This trend can partially be explained by a limitation in the lead time approximation which in the case of relatively long lead time becomes even more evident. Since the market investigated in this master thesis involves retailers replenishing from Johannesburg, the deviation is likely to be smaller in other markets with shorter lead time controlled by SML. The replenishment lead time to Johannesburg from the CDC in Gent is one of the longest due to the geographical distance.

Given the result from the numerical study, and the limited number of items tested, the model seems to perform well in achieving target fill rates while simultaneously reducing the average inventory in the system. Given this result, we can expect to see similar results for a wider sample size.

8.1 Future research

Modeling stochastic processes necessitates making assumptions and approximations to render them computationally feasible. In this particular model, a logarithmic distribution has been chosen to approximate the customer order sizes. The decision to approximate the demand as logarithmic stems from insufficient data and time constraints. Nonetheless, for effective implementation of the model, additional data or improved approximations for the demand are imperative.

Incorporating emergency shipment and return mechanisms into the model would better represent the real dynamics of inventory distribution. However, the absence of established criteria for initiating these actions poses a challenge. Without a robust framework defining when such events occur, integrating these functions becomes impractical.

Additionally, extending the model into three echelons, by including the central warehouse distributing to the regional warehouse, would be interesting to see how the inventory is allocated. Currently, we assume an infinite stock supplying the RDC, whereas in practice, this is not the case.

References

Alvarez, E., Heijden, M. & Zijm, W. (2015). Service differentiation in spare parts supply through dedicated stocks. *Annals of Operations Research*, Vol. 231, pp. 283–303.

Andersson, J., Axsäter, S. & Marklund, J. (1998). Decentralized multi echelon inventory control. *Production and Operations Management*, Vol.7, pp. 370–386.

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

Axsäter, S. (2006). Inventory Control, 2:th ed. *Springer Science + Business Media, LCC.*

Axsäter, S., Kleijn, M. & de Kok, T. G. (2004). Stock Rationing in a Continuous Review Two-Echelon Inventory Model. *Annals of Operations Research*, Vol. 126, pp. 177–194.

Axsäter, S., Olsson, F. & Tydesjö, P. (2007). Heuristics for Handling Direct Upstream Demand in Two-Echelon Distribution Inventory Systems. *International Journal of Production Economics*, Vol. 108, pp. 266–270.

Berling, P., Johansson, L. & Marklund, J. (2023). Controlling inventories in omni/multi-channel distribution systems with variable customer order-sizes. *Omega*, Vol. 230, pp. 515–526.

Berling, P. & Marklund, J. (2006). Heuristic Coordination of Decentralized Inventory Systems Using Induced Backorder Costs. *Production and Operations Management*, Vol. 15, pp. 294–310.

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

Berling, P. & Marklund, J. (2014). Multi-echelon inventory control: an adjusted normal demand model for implementation in practice. *International Journal of Production Research*, Vol. 52, pp. 3331–3347.

Dekker, R., Kleijn, M. J. & de Rooij, P. J. (1998). A spare parts stocking policy based on equipment criticality. *International Journal of Production Economics*, Vol. 56–57, pp. 67–77.

Harris, F. W. (1913). How Many Parts to Make at Once. Factory, *The Magazine of Management,* Vol. 10, pp- 135-136.

Hillier, S. & G. J. Lieberman. (2010) Introduction to operations research, 9:th edition. *McGraw-Hill, The McGraw-Hill Companies.*

Höst, M., B. Regnell, and P. Runeson. (2006) Att genomföra examensarbete. *Studentlitteratur.*

Laguna, M. & Marklund, J. (2018). Business Process Modeling, Simulation and Design. *Taylor and Francis Group*.

Nahmias, N. & Demmy, W. S. (1981). Operating characteristics of an inventory system with rationing. *Management Science*, Vol. 27, pp. 1236–1245.

Schulte, B. & Pibernik, R. (2016). Service differentiation in a single-period inventory model with numerous customer classes. *OR Spectrum*, Vol. 38, pp.921–948.

Schulte, B. & Pibernik, R. (2017). Profitability of Service-Level-Based Price Differentiation with Inventory Rationing. *Production & Operations Management*, Vol. 26, pp. 903–923.

Sherbrooke, C. C. (1968). Metric: A multi-echelon technique for recoverable item control. *Operations research*, Vol. 16, pp. 122–141.

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

Veinott, A. F. Jr. (1965). Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes. *Operations Research*, Vol. 13, pp. 761–78.

Volo Group. (2024a). Company presentation. Available at: https://www.volvogroup.com/content/dam/volvo-group/markets/global/en-en/company-presentation/Volvo_Group_Company_Presentation_2306.pdf (Accessed: 2024-01-09).

Volvo. (2024b) *Volvo Group intranät [internal material]*.

Volvo. (2024c) *Volvo SML presentation [internal material]*.

Appendix A: Interview guide

Person introduction
What is your position at Volvo?
What types of tasks do you work with on a regular basis?

General
Could you walk us through the Volvo organization, how does Volvo spare parts fit in this?
Do you have any organizational tree that we could have access to?
How is the system integrated with other business systems, such as ERP?
What is the procedure if the systems is not the same in different business units?
What does a project at SML normally look like?
Who are your end customers?
What is significant for these customers?
How is the contracts with the customers normally designed? Do you for example have KPIs such as service levels incorporated in the contracts?
What key performance indicators (KPIs) do you use to evaluate the efficiency of your distribution and inventory management system?
How does the fact that you are working with spare parts impact your way of structuring?
What challenges or pain points have you encountered with your current system?

General Distribution Network
What does your distribution network look like?
- Different depending on market/geographical location?
- What types of distribution channels do you use? (market/geographical dependent?)

Do you ship partial deliveries or only full orders?
Is your system generally a divergent system? (other words: retailers only ordering from one RDC)
Do you have a map of the distribution network?
Do you use vendor managed inventory agreements?

Do you use any differentiation in priority to the different customers/markets?
Which level of service differentiation do you use today?
How do you accommodate for the different requirements?
Translateral shipment between dealers?
What strengths do you see in your current network?
What weaknesses do you see?

Inventory Management System
How do you segment products?
Are there some segments particularly hard to manage? (sizes, available space etc).
What kind of inventory policy do you use?
Do you have the same policy everywhere? Dependent on specific market?
Do you use continuous or periodic review? If periodic, which periodicity?
What is your current method of determining parameters for inventory control?

- Can you walk us through the process?
- Why have you chosen this way of optimizing? Strengths/Weaknesses?

At which frequency are these updated?
How is the current inventory management system performing?
Do you use deterministic or stochastic calculations?
Which assumptions and approximations are made?
How do you track and manage inventory levels?
How are customer orders processed and tracked from placement to delivery?
Is there a First Come First Serve (FCFS) policy or priority depending on dealer?
What measures are in place to handle out-of-stock situations? are “missed orders” back-ordered or counted as “lost sales”?

- Is it different depending on dealer?

How are these costs calculated?
Is slow and lumpy a good description of the demand for your products?

- Do you have spare parts exhibiting other demand patterns? (e.g. normal distribution, high quantities etc)

- Are there different distributions depending on demand patterns used?

Any restrictions regarding order quantities from dealers?

Available data

What type of data is available for the different items/markets in general?

How is demand data organized at different aggregate levels? For example, is it stored at the SKU-level, SKU at warehouse-location level, or any other specific level?

What is the time interval for storing demand data? Is it recorded on a weekly, monthly, or daily basis? Additionally, can we retrieve information on past deliveries, including order times and quantities?

Can we retrieve past inventory policy parameters for analysis and reference?

Is historical stock-level data accessible, allowing us to track changes over time?

Is information available regarding delivery times between various warehouse locations, encompassing both transportation times and the actual delivery times (i.e., the duration between order placement and delivery)?

Appendix B: List of notation

R_i - Reorder point at retailer i

R_0 - Reorder point at central warehouse

R^* - Optimal reorder point

Q_i - Order quantity at retailer i

Q_0 - Order quantity at central warehouse

$Q_{i,n}$ - Order quantity at retailer i in normalized system

λ - Arrival rate (mean within time period t)

f_j - The probability of a certain demand size of j by a single customer

f_j^k - The probability of k customers having a total demand of j

$D(t)$ - Stochastic demand during the time interval t

D_i - Stochastic demand at installation i

O_i - Discrete stochastic order size at installation i

μ - Expected mean demand per time unit t

μ′ - Estimation of the expected mean demand during the time unit t

σ^2 - Variance of the demand

σ - Standard deviation of the demand

σ′ - Estimation of the standard deviation of the demand

IL_i - Inventory level at installation i

IP_i - Inventory position at installation i

L - Lead time to upstream installation

$\overline{L}_i$ - Estimated expected lead time at installation i

$\widehat{L}_i$ - Expected lead time for virtual retailer with delivery delay

l_i - Transportation time from the central warehouse to the regular retailer i

f - Density function

F - Distribution function

$g(x)$ - The gamma distribution function

$\Gamma(r)$ - The gamma function

h_i - Holding cost at installation i

b_i - Backorder cost at installation i

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

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

β_i - Induced backorder cost at installation i

β^* - Optimal induced backorder cost

$\beta_{i,n}$ - Induced backorder cost at installation i in normalized system

p_i - Shortage cost per unit and time unit

$p_{i,n}$ - Shortage cost per unit and time unit in normalized system

$\widehat{p}_i$ - Shortage cost per unit an time unit for the customer demand at retailer i

S_1 - Probability of no stockout per order cycle

S_2 - Fill rate

S_3 - Ready rate

$G(x)$ - Loss function

ϕ - Density function for normal distribution N(0,1)

Φ - Probability function for normal distribution N(0,1)

s_i - Base stock level

N - Number of regular retailers

M - Number of virtual retailers

γ_i - Fill rate at retailer i (i = 1,...,N)

γ_i^* - Target fill rate for the demand at retailer i (i = 1,...,N)

Appendix C: Induced backorders - Normalization of system

	Original system parameters	**Normalized system parameters**
Retailer order quantity	Q_i	$Q_{i,n} = \frac{100Q_i}{\mu_i l_i}$
Central warehouse order quantity	Q_0	$Q_{0,n} = Q_0$
Retailer holding cost per unit and time unit	h_i	$h_i = 1$
Central warehouse holding cost per unit and time unit	h_0	$h_{0,n} = \frac{h_0}{h_i}$
Retailer shortage cost	p_i	$p_{i,n} = \frac{p_i}{h_i}$
Central warehouse lead-time	L_0	$L_{0,n} = \frac{L_0}{l_i}$
Retailer transportation time	l_i	$l_{i,n} = 1$
Expected demand per time unit at retailer i	μ_i	$\mu_{i,n} = 100$
Standard deviation of demand per unit and time unit at retailer i	σ_i	$\sigma_i, n = \frac{100\sigma_i}{\mu_i\sqrt{l_i}}$
Induced backorder cost	$\beta_i = \beta_{i,n} \cdot h_i$	$\beta_{i,n}$

$$\beta_i = h_i \cdot g(Q_{i,n}, p_{i,n}) \cdot \sigma_{i,n}^{k(Q_{i,n}, p_{i,n})}, \ For\ i = 1, 2... N$$

$$g(Q_{i,n}, p_{i,n}) = min[g_a \cdot (Q_{i,n})^{g_b}, G]$$

$$g_a = min[0.015p_{i,n}, max(\frac{0.65}{\sqrt{p_{i,n}}}, 0.05)]$$

$$g_b = max[-1.2, -2p_{i,n}^{-0.25}]$$

$$G = min[0.015, 0.005p_{i,n}^{0.2}]$$

$$k(Q_{i,n}, p_{i,n}) = max[1, min(k_a \cdot Q_{i,n}^{k_b}, K)]$$

$$k_a = max[0.7, min(0.9, 0.6p_{i,n}^{0.075})]$$

$$k_b = min(0.2, 0.4p_{i,n}^{-0.35})$$

$$K = max(1.3, min(2, 2.5p_{i,n}^{-0.15}))$$

Appendix D: Test data multiple channels

Test series	**Mean demand and standard deviation for regular retailers per time unit. (μ, σ)**	**Mean demand and standard deviation for virtual retailers per time unit. (μ, σ)**	**Demand in each channel (%)**	**Target FR**	**Difference from target FR (pp)**
1	0,0555; 1	VR1: 0,75; 1 VR2: 0,75; 1	Regular: 25% VR1: 37,5% VR2: 37,5%	VR1: 80% VR2: 95%	VR1: +7,3 VR2: 0
	0,0555; 1	VR1: 0,75; 1 VR2: 0,75; 1	Regular: 25% VR1: 37,5% VR2: 37,5%	VR1: 90% VR2: 98%	VR1: +3,4 VR2: - 2,1
2	0,0555; 1	VR1:0,75; 5 VR2:0,75; 5	Regular: 25% VR1: 37,5% VR2: 37,5%	VR1: 80% VR2: 95%	VR1: +4,7 VR2: +0,5
	0,0555; 1	VR1:0,75; 5 VR2:0,75; 5	Regular: 25% VR1: 37,5% VR2: 37,5%	VR1: 90% VR2: 98%	VR1: +0,9 VR2: +0,5
3	0,0555; 1	VR1:1,2; 5 VR2:0,3; 5	Regular: 25% VR1: 60% VR2: 15%	VR1: 80% VR2: 95%	VR1: +6 VR2: +1,3
	0,0555; 1	VR1:1,2; 5 VR2:0,3; 5	Regular: 25% VR1: 60% VR2: 15%	VR1: 90% VR2: 98%	VR1: -1,1 VR2: +0,6
4	0,0555; 1	VR1:0,3; 5 VR2:1,2; 5	Regular: 25% VR1: 15% VR2: 60%	VR1: 80% VR2: 95%	VR1: +10,5 VR2: +1,2
	0,0555; 1	VR1:0,3; 5 VR2:1,2; 5	Regular: 25% VR1: 15% VR2: 60%	VR1: 90% VR2: 98%	VR1: +5,5 VR2: -0,4
5	0,1667; 1	VR1:0,25; 5 VR2:0,25; 5	Regular: 75% VR1: 12,5% VR2: 12,5%	VR1: 80% VR2: 95%	VR1: -0,7 VR2: +0,8
	0,1667; 1	VR1:0,25; 5	Regular: 75%	VR1: 90%	VR1: -2,0

		VR2:0,25; 5	VR1: 12,5% VR2: 12,5%	VR2: 98%	VR2: +0,9

Appendix E: Active retailer for items

ITEM	Info	System parameters Q_{RDC}; μ^{tot}; σ^{tot}; % demand in VR channels	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	VR1	VR2
1	Initial 10	118; 2,58; 2,77; 27%	x	x	x	x	x	x	x	x	x		x	x	x		x
2	Initial 10	98; 1,23; 3,52; 22%	x			x		x	x	x	x	x	x	x			x
3	Initial 10	341; 11,47; 11,6; 39%	x	x	x	x	x	x	x	x	x	x	x	x		x	x
4	Initial 10	149;2,38;3,33; 55%	x			x	x	x		x	x	x	x	x			x
5	Initial 10	40; 0,73; 1,27 24%				x	x	x	x	x	x	x	x	x			x
6	Initial 10	27; 2,11; 2,34; 14%				x	x	x	x	x	x	x	x	x	x		x
7	Initial 10	115; 2,06; 2,65; 23%	x	x		x	x	x	x	x	x	x	x	x			x
8	Initial 10	10; 0,76; 1,65; 38%	x		x	x		x	x	x	x	x	x	x			x
9	Initial 10	72; 9,5; 6,5; 25%		x	x	x	x	x	x	x	x	x	x	x	x	x	x
10	Initial 10	116; 2,46; 4,42; 46%	x		x	x	x	x	x	x	x	x	x	x	x		x
11	Price class 9	18; 2,17; 2,5; 6%		x	x		x	x	x	x	x	x		x		x	
12	Price class 9	11; 1,4; 1,4; 23%				x	x	x	x	x	x		x	x	x	x	x

13	Price class 6	65; 8,37; 5,2; 33%	x	x	x	x	x	x	x	x	x	x		x		x	x
14	Price class 3	68; 2,17; 2,47; 51%	x		x	x		x		x		x	x	x	x	x	x
15	Price class 1	437; 15,97; 20,96; 24%	x		x	x	x	x	x	x	x	x	x	x	x	x	x