Inventory Management and Demand Fulfilment in Omni-channel Retail

The Role of the Store

Joost Goedhart

Propositions

1. An omni-channel retailer must use both stores and warehouse inventories for the optimal fulfilment of online demand.
(this thesis)

2. A high online return rate decreases the offline sales of an omni-channel retailer.
(this thesis)

3. Obtaining a PhD is like applying reinforcement learning.

4. Artificial intelligence will guide but not dominate our decisions.

5. Sometimes having no opinion is better than having one.

6. Snappy PhD propositions are in conflict with the nuance of science.

Propositions belonging to the thesis, entitled

Inventory Management and Demand Fulfilment in Omni-channel Retail: The Role of the Store

Joost Goedhart
Wageningen, 9th June 2023

Inventory Management and Demand Fulfilment in Omni-channel Retail: The Role of the Store

Joost Goedhart

Thesis committee

Promotors:
Prof. Dr S.L.J.M. de Leeuw
Professor of Operations Research and Logistics
Wageningen University & Research

Dr R. Akkerman
Associate Professor, Operations Research and Logistics
Wageningen University & Research

Co-promotor:
Dr R. Haijema
Associate Professor, Operations Research and Logistics
Wageningen University & Research

Other members:
Dr H.W.I. van Herpen, Wageningen University & Research
Prof. Dr S. Bhulai, Vrije Universiteit Amsterdam
Prof. Dr A. Huchzermeier, WHU—Otto Beisheim School of Management, Vallendar, Germany
Dr R.P. Rooderkerk, Erasmus University Rotterdam

This research was conducted under the auspices of the Wageningen School of Social Sciences (WASS)

Inventory Management and Demand Fulfilment in Omni-channel Retail: The Role of the Store

Joost Goedhart

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 9 June
at 4 p.m. in the Omnia Auditorium.

Joost Goedhart
Inventory Management and Demand Fulfilment in Omni-channel Retail:
The Role of the Store
218 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)
With references, with summaries in English and Dutch

ISBN 978-94-6447-592-0
DOI https://doi.org/10.18174/587326

Contents

Chapter 1

Introduction

1.1 Omni-channel Retailing

In recent times, retailing has changed drastically. In the past, customers could only visit brick-and-mortar stores for the fulfilment of their demand for goods; nowadays, customers are presented with different options to fulfil their demands (Bell et al., 2014). With the advancement of the internet, customers can buy products at online retailers and have them shipped to their homes (Marhamat, 2022). Nonetheless, in only a few decades, online retailers have started to pose a significant threat to traditional retailers. For example, in the US, revenue from online sales grew from 145 billion dollars in 2009 to 579 billion dollars in 2019[1] (U.S. Department of Commerce, 2011, 2021). Customers shifted from shopping in-store to online, resulting in a decline in brick-and-mortar sales. Convenience, efficiency, and lower prices are some of the reasons for customers to shop online (Perea y Monsuwé et al., 2004). This development also resulted in a decline of brick-and-mortar stores, where mostly small retailers shut down as they could not compete against the large online retailers (Bell et al., 2018b). In same cases, this even led to a decline in shopping streets in the city centers (Carmona, 2022).

To compete against the threat of online retailers, retailers with brick-and-mortar stores also started adopting online retailing in their portfolio (Deloitte, 2015). They continued to sell products in their brick-and-mortar stores, but also offered products via web stores. Often, these two sales channels were operated individually, where the price and assortment of the channels differed (Verhoef et al., 2015). Additionally, physical assets such as distribution centres, inventories, and delivery vehicles were often operated individually instead of being used to support each other (Hübner et al., 2016a). This strategy is often referred to as multi-channel retailing (Brynjolfsson et al., 2013).

Customers shopping at a multi-channel retailer thus have different shopping channels available. This results in that a customer might switch between channels during their shopping activities, as described in Bijmolt et al. (2021). An initial demand can be established by visiting a store, but the information finding is done via the online channel (Goraya et al., 2022). Finally, the customer might still be buying the product in the store to receive immediate gratification for obtaining the product.

This new development of customers switching between channels during their cus-

[1]In the following years, the growth almost even doubled, albeit skewed as the lock-downs during the COVID-19 pandemic resulted in brick-and-mortar stores temporary closing and customers fulfilling their needs online.

tomer journey highlighted a new problem for retailers: customers are dissatisfied if the information, assortment, or prices differ between the channels. They consider a retailer to be one entity and therefore expect that there is no difference between the brick-and-mortar store and the web store (Abrudan et al., 2020). Therefore, retailers are now often integrating their online channel with their brick-and-mortar channel, resulting in what is referred to as omni-channel retailing. Verhoef et al. (2015) gives a formal definition of omni-channel management:

"the synergetic management of the numerous available channels and customer touchpoints, in such a way that the customer experience across channels and the performance over channels is optimized"

To optimise the customer's experience, there are then less differences between channels, as customers prefer to have the same information, assortment, and price in all channels (Abrudan et al., 2020). The omni-channel retailer needs to coordinate all channels to ensure this seamless experience. From a marketing perspective, this would mean that the customer's perceived interaction is with the retailer, not the channel it visits (Piotrowicz & Cuthbertson, 2014). This requires careful marketing from the retailer's perspective to ensure that during the customer's journey, a uniform experience is provided to optimise the customer's perceived value with the retailer.

However, from a retailer's perspective, this does not include the retailer's challenges and opportunities as it solely focuses on optimising customer experience. Ailawadi & Farris (2017) notes that omni-channel marketing should also be about the retailer's operations. Cui et al. (2021) proposed a novel goal for omni-channel marketing, which involves the optimisation of the customer experience for the benefit of both the retailer and their customers.

Optimising the customer experience is hindered by the increasing customer demand, which extends further than the need to fulfil demand. Customers prefer a shopping experience in which they can expect a wide range of products for low prices, fulfilment via their preferred method, and lenient return policies (McKinsey, 2021a). To fulfil these customers' needs, retailers nowadays have a web page, mobile application, in-store information screens, and other touch-points where customers and retailers meet. All these developments are aimed to merge the digital and physical channels of the retailer to offer a seamless customer shopping experience (Rooderkerk & Kök, 2019).

Although this flexibility improves a customer's shopping experience (e.g. Herhausen et al., 2015), a retailer may need to change the inventory and assortment decisions to different channels in order to satisfy the customer's needs. Where the marketing provides a seamless experience to the customer, the retail operations have to make it happen (Jasin et al., 2019). The retail operations focus on fulfilling the needs of the customer in a cost-efficient manner. With omni-channel retailing, this interaction is increasingly complex as the customer has access to more product options, delivery methods, and return possibilities. Often, retailers first start as omni-channel retailers from a marketing perspective but the operations of the different channels are separated. Only in a later stage, the operations are integrated (Rooderkerk et al., 2023). EY (2015) mentions that when a retailer becomes an omni-channel retailer they need to reshape their supply chain as it is not fit for the new operations.

1.2 Omni-channel Operations Challenges

Melacini et al. (2018) identifies three different types of omni-channel retail operations challenges that companies experience when implementing an omni-channel retailing strategy: (i) network design, (ii) inventory management, and (iii) demand fulfilment. The main issue with retail operations in an omni-channel context is the integration and coordination of the different channels in which suddenly the brick-and-mortar store and online channel have to support each-other (Piotrowicz & Cuthbertson, 2014).

One of the challenges with network design is integrating the online channel into the current supply chain, the retailer has to think about the level of integration in the physical network. The retailer has the option to introduce an online fulfilment centre (OFC) for online orders, which can often offer a wider range of products and has more storage space than stores (EY, 2015). However, if this results in the online channel becoming separated from the store there will be less synergy between the channels and is not considered as omni-channel retail operations (Rooderkerk et al., 2023). Integrating both channels would result in leveraging the store network for the online channel, which requires careful coordination and can be complex.

The challenge with inventory management in omni-channel retailing is that the retailer has the decide how to use their inventory for the different channels. Stores can now be integrated with the online channel and therefore their inventory can thus serve for fulfilling online demand, resulting in a demand pooling effect (ENC, 2016). This can reduce uncertainty in demand and therefore the total inventory levels across the network (Hovelaque et al., 2007). However, the store has often a fixed assortment as shelf spaces are limited, while an OFC has almost limitless shelf spaces (Noble et al., 2005). This can create conflict between channels, as the online channel might offer more products than in-store. Thus the retailer has to carefully think about what products to offer in each channel to ensure the overall customer experience is optimised. A potential solution to also offer unlimited assortment in the offline channel is having tablets available in-store, which offer the products for delivery to the store or the customer's home (Piotrowicz & Cuthbertson, 2014). Another complexity that arises is the online ordered products might be returned to the store, as omni-channel retailing allows for cross-channel returns. This might disrupt the inventory management of the store and should be accounted for (Bernon et al., 2016).

The challenge with demand fulfilment is that the retailer has to decide how to enable different fulfilment processes for the customer. It can offer buy-online-pick-up-in-store (BOPS), which means that the customer can place an order online and collect it in a store (Song et al., 2020). This is advantageous for the customer as it can collect a product in a store without the risk of encountering a stock-out. When a BOPS order is fulfilled via the OFC (referred to as ship-to-store strategy), instead of from store inventory, it allows the retailer to offer more products than the store assortment as the OFC can hold more products (Melacini et al., 2018). However, using in-store inventory would avoid shipping costs, as the product is already located at the pick-up point (Gao & Su, 2017a). Another fulfilment strategy for the retailer is to decide to ship online ordered products from the store to the customer's home (referred to as ship-from-store strategy). An advantage of this strategy is that the delivery cost can be reduced as customers are often located closer to stores than to OFCs (Deloitte, 2015). Additionally, the product can be delivered rapidly due to the short delivery distance, in some cases allowing for two-hour delivery (McKinsey, 2020).

The three operations challenges introduced above that companies experience when implementing omni-channel retailing are relevant for both small retailers operating a single store and large retailers operating multiple stores. The COVID-19 pandemic lockdowns recently caused retailers to not have in-store customers, resulting in them opening web shops to sell their goods (PWC, 2020). This mostly resulted in smaller retailers with only one brick-and-mortar store leveraging their store for the online channel. After the initial lockdowns, they often continued to use their web shops. This also meant that they have to investigate how to use their current stores for both channels, where they have to integrate assortment and inventory. The challenge for larger retailers is similar to those operating a single store, but they have an additional decision of which store will be used for supporting the online channel (Govindarajan et al., 2021). Also, larger retailers might have to decide if they want to open an online fulfilment centre and decide its role in the fulfilment of online orders. The retailer has to allocate the online order to a fulfilment location, taking into account costs and potential future demand (Arslan et al., 2021). For proper integration of the online and offline channels, retailers have to redefine the role of the store and its operations.

1.3 Redefining the Role of the Store

As mentioned above, the store has become more than a place for customers to purchase a product with omni-channel retailing. In fact, it has become a place to gain information about the product, order a product, return a product, or receive an online ordered product (Hübner et al., 2022). As a result, a retailer may need to reconsider the role of their brick-and-mortar store. Instead of just a place to shop and buy, a store may be a valuable asset if it can physically support the online channel. Although a challenging task, it also creates new opportunities for operational synergies between offline and online channels (Ishfaq et al., 2016).

Hübner et al. (2022) presents an overview of the current state of the literature on store operations in an omni-channel retail context. They classify the literature into three groups: (i) demand forecasting, (ii) network design of fulfilment locations, and (iii) assortment and inventory management.

Demand forecasting relates to the generation of accurate forecasts, but also to analysing how customers switch between channels. Studies show that integrating demand forecasting and operations results in lower costs and a better match between demand and supply (Pereira & Frazzon, 2021). However, most studies related to demand forecasting focus on BOPS while other omni-channel fulfilment options, such as ship-to-store or buy-online-return-to-store, would be of interest as well (Omar et al., 2022).

Network design relates to defining which locations should be used for order fulfilment and which order should be fulfilled from what location. Most literature focuses on either the strategic decision on what type of omni-channel fulfilment options to adopt (e.g. Bayram & Cesaret, 2021; Saha & Bhattacharya, 2021) or the operational decision of the online order allocation (e.g. Andrews et al., 2019; Govindarajan et al., 2021) but few integrate these decisions.

Assortment and inventory management relates to the selection of products available for the customer and to the inventory replenishment decisions. The literature often focuses on the ordering decision (e.g. Gao & Su, 2017b; Radhi & Zhang, 2019), but few papers focus on how these decisions are influenced by the potential returns of products sold online. Furthermore, studies often focus on a single store or omni-channel fulfilment option, which is of interest to some retailers but not all.

1.4 Inventory Management and Demand Fulfilment

The above-mentioned dynamics show that omni-channel retailers have new operational challenges in their stores. The inventory management needs to be revised, as it now not only serves in-store customers. Taking into account these new customer needs can be challenging, as it not only imposes extra demand and uncertainty through potential returns, but it also results in extra decision-making. The omni-channel retailer has to rethink the inventory management due to these new challenges.

Using the store inventory for the fulfilment of online orders is one of the earliest and most easily adopted roles that the stores can fulfil to support the online channel (Bendoly et al., 2007). Stores can be used to have customers pick up the products they ordered online, or online orders can be picked and shipped from stores to the customer's home (Baird & Kilcourse, 2011).

Using the store inventory to support the online channel has many advantages. As online and in-store demand is pooled at one location, demand uncertainty is pooled, resulting in lower total safety stocks. This can reduce overall inventory cost for the retailer (Jalilipour Alishah et al., 2015). Additionally, higher inventory turnover rates due to the demand pooling effect can reduce the number of leftovers at the end of the sales period (Bayram & Cesaret, 2021). Finally, stores are often located closer to the customer than online fulfilment centres, thus environmental and economic benefits can be achieved for the delivery of online orders (Dethlefs et al., 2022).

However, using store inventory to fulfil online demand also brings challenges. Conflicts between channels might occur in different ways. A seemingly available product could be bought by customers in both channels, resulting in order cancellations and related customer dissatisfaction (Nguyen et al., 2018). Furthermore, a product ordered online might need to be picked from store shelves, resulting in order pickers working in the same space as customers shopping. This may negatively influence the customer's in-store shopping experience (Ishfaq & Bajwa, 2019). In addition, store shelves are not designed for picking orders thus the process of order picking is not efficient (Hübner et al., 2016c). Finally, when a retailer has multiple stores available for fulfilling online orders, a decision needs to be made about which store inventory to use (Jiu, 2022).

The store can not only be used to fulfil online orders, but also as a location for customers to return products ordered online. As customers are ordering more products online, the number of products being returned also has grown (McKinsey, 2021b). This return flow has become a significant issue, as sometimes almost half of the online ordered products are being returned (de Leeuw et al., 2016). Often, these returned products can be sold again, as retailers have the policy that returned products should be unused and labelled. Therefore returned products should be accounted for in inventory management practices (Radhi & Zhang, 2019).

The advantages of having customers return products in-store are that the retailer can advise a replacement (think of another clothing size if it did not fit) or steer them towards buying another product (Wollenburg et al., 2018). In addition, store personnel can act as a gateway, ensuring that the returned product is in valid condition to be returned after inspection (de Leeuw et al., 2016). Furthermore, for the retailer, it can be more profitable to have the customer return the product in stores, as often the shipping cost of returning the product is incurred by the retailer (Deloitte, 2020). Finally, returning a product via the store has the advantage that it can quickly be added back to the store inventory, therefore not losing the value of the product (Hübner et al., 2015).

However, one of the most challenging aspects of product returns is forecasting them (Hübner et al., 2022). Not only is it uncertain whether a product is returned, but it is also uncertain when a product is being returned (Muir et al., 2019). This increases the complexity of inventory management, as not taking into account these possible return flows might result in excessive stock. Further complications might arise when the original online order that is being returned in-store was fulfilled via an online fulfilment centre. These cross-channel returns might result in unbalanced inventories across the retailer's network (Dijkstra et al., 2019).

Although research has been done on retail operations in an omni-channel retailing context, many challenges still remain unaddressed for the store inventory management. Those studies that did investigate store inventory management typically focused on a single operational decision (e.g., stock ordering or order allocation), not taking into account the interactions between these decisions. Furthermore, studies typically consider that products are sold in a limited selling period, whereas retailers might also sell products that have regular replenishment schedules. Lastly, most research related to store operations and omni-channel retailing focuses on shipping online orders from the store (i.e., ship-from-store), not taking into account other fulfilment options such as ship-to-store, BOPS, or buy-online-return-to-store.

1.5 Research Outline and Aims

This thesis examines the role of a brick-and-mortar store in omni-channel retail in general, as well as some of the underlying operational decision processes. We specifically investigate two different types of settings for an omni-channel retailer. We investi-

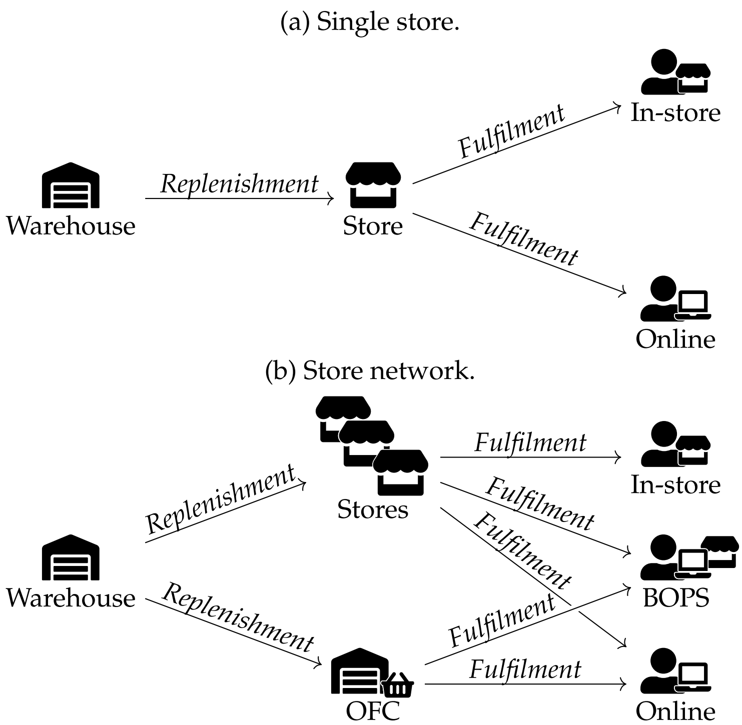


Figure 1.1: Visualisation of the flow of goods for a single store (a) and a store network (b) for an omni-channel retailers studied in this thesis (OFC: online fulfilment centre, BOPS: buy-online-pick-up-in-store).

gate the setting in which the omni-channel retailer has a single store or when multiple stores and a potential OFC are available. Figure 1.1 shows the flow of goods from the omni-channel retailer to the different types of customers that we study in this thesis. Figure 1.1(a) depicts the supply chain of an omni-channel retailer with a single store in which the store inventory is used to fulfil both in-store demand and online demand, which is investigated in Chapter 2 and 3. Figure 1.1(b) shows the flow of goods from the stores or OFC to the different types of customers, which now also includes BOPS customers. In this setting, the omni-channel retailer who has multiple locations from which to fulfil online demand, Chapter 4 and 5 investigate the inventory management in this setting.

1.5.1 Omni-channel retailers single store operations

Chapter 2 is concerned with how a brick-and-mortar store inventory can be utilised to fulfil online demand. More specifically, in this research we focus on how to allocate inventory to the different demand channels via a rationing decision. We include a replenishment decision with both decisions happening in different time periods (e.g., daily and weekly). We develop a heuristics that performs near-optimal and reduces the computational complexity.

Chapter 3 focuses on the inventory management of a brick-and-mortar store that also receives returned products from online sales. The uncertainty of the returns complicates the retailers' operations. As we explicitly model multi-period and sales-dependent returns, the studied problem easily becomes intractable. Therefore, we construct a Deep Reinforcement Learning algorithm to solve the studied problem.

1.5.2 Omni-channel retailers network operations

Chapter 4 addresses which stores to fulfil online orders from when the retailer has multiple stores available. When allocating an online order to a store, the retailer has to take into account several factors such as inventory levels and future demand. Current rules are often myopic and result in imbalanced inventory across stores. We develop a policy that keeps inventory levels across the stores more balanced by incorporating these factors.

Chapter 5 discusses which fulfilment strategies an omni-channel retailer can adopt. Customers can order and collect their products via different channels, and the retailer can source their products from different locations. The retailer has to decide which locations to include for fulfilling demand and then allocating the orders to the different locations. Therefore, we investigate four typical fulfilment strategies used in practice with a focus on the inventory management of the stores and the OFC.

Finally, in Chapter 6, we summarise the main findings of the thesis in a general conclusion. We integrate the findings of the different studies and discuss the implications and limitations of the thesis. The chapter ends with suggestions for future research.

Chapter 2

Inventory Rationing and Replenishment for an Omni-channel Retailer

This chapter is published as:

Goedhart, J., Haijema, R., & Akkerman, R. (2022). Inventory rationing and replenishment for an omni-channel retailer. *Computers & Operations Research, 140*, 105647

Abstract

The growth of omni-channel retailing resulted in many new challenges for retailers, especially in relation to the replenishment and allocation of inventories for the different channels. In this paper we consider a brick-and-mortar store that uses its inventory to fulfil both in-store demand as well as online orders. In addition to deciding on replenishment quantities, such a retailer also has to decide how to ration its inventory across the channels. Practically, rationing inventory relates to storing part of the inventory in the backroom to satisfy online demand. The rationing process occurs regularly (e.g. daily) whereas inventory replenishment typically occurs less regular. To analyse this decision process, we model the rationing and ordering decisions as a Markov Decision Problem that maximises the expected profit. Based on the structure of the optimal policies, we determine heuristics that near-optimal results and scale well to retailers with many products.

2.1 Introduction

The increased competition of online shopping has put pressure on brick-and-mortar stores. Online shopping has made it effortless for customers to satisfy their demands from home. To compete against online retailers, physical store retailers are increasingly adopting online shopping channels in their channel portfolio. This integration of different sales channels is referred to as omni-channel retailing (Verhoef et al., 2015). The goal of omni-channel retailing is to provide customers with a seamless shopping experience and enhance customer loyalty and satisfaction.

When an omni-channel retailer adopts new channels, it needs to reconsider its inventory policies (Jalilipour Alishah et al., 2015). Retailers for instance have to decide whether or not to integrate inventories of different channels. Retailers that add an online channel often choose to use store inventory to satisfy online demand (ENC, 2016), as this has the lowest initial investment (Fernie & Sparks, 2004). The integration of online sales with offline sales is an ongoing discussion in the field of retail operations and is referred to as bricks-and-clicks (Agatz et al., 2008). Especially smaller retailers are adopting this concept, in which their store essentially become a distribution centre for their online order fulfilment (Mou et al., 2018). In this way, retailers are able to leverage their brick-and-mortar store with online sales.

By using store inventory for the fulfilment of online orders, several advantages can be achieved. First, since offline inventory is also used to serve online customers, uncertainty in demand can be reduced, lowering the total inventory. Second, higher inventory turnover rates result in a decrease in left-overs at the end-of-sales periods (Bendoly, 2004; Bayram & Cesaret, 2021). Furthermore, stores are often located closer to customers than distribution centres, thus environmental and economic benefits for the delivery of online orders can be achieved (Jalilipour Alishah et al., 2015).

However, using stores to fulfil the demand of online customers also has disadvantages. Using store personnel for picking orders in-store might influence the customer shopping experience and can be a high expense for the retailer (Baird & Kilcourse, 2011; Ishfaq & Bajwa, 2019). Furthermore, store inventory needs to be monitored more closely to ensure online orders can be satisfied. If online demand occurs, this means picking the order from the store shelves and updating the inventory level on the website. If a customer buys the product in-store, this also needs to be coordinated with the online channel to ensure that there is no conflict in which an online customer tries to buy the same product digitally. To mitigate these problems retailers often choose to satisfy online demand from the remaining store inventory at the end of the day when customers have left. However, stock-outs might occur resulting in decreased customer loyalty and satisfaction as online demand cannot be satisfied (Nguyen et al., 2018).

To mitigate the negative effects of using a store as a fulfilment centre for online orders, managerial studies (e.g. Hobkirk, 2015; ENC, 2016) have suggested to reserve a certain amount of in-store inventory to satisfy online demand. In practice, this relates to storing the inventory for the online demand in the backroom (Aastrup & Kotzab, 2010). By doing so, the retailer does not need to continuously coordinate the two sales channels, preventing sales of products that turn out to actually be unavailable. Furthermore, there is no need to pick online orders at the moment they arrive. The retailer can inform their online customers immediately whether a product is available in their respective channel. Also for the in-store channel, the retailer needs to know how many items to display at the store, as a product that is no longer on display may generate no demand. Managing the store inventory happens at a daily basis, thus unbalanced inventory among channels can be reduced by regular rebalancing of the inventory. When retailers ration their inventory across their physical and digital channels in this way, they can balance the satisfaction of demand from both channels. This avoids the disadvantages outlined above, while creating a new one: when a stock-out happens in one of the channels, a lost sale occurs even though the other channel might have inventory left.

If a retailer rations its inventory, a trade-off therefore exist between serving in-store customers and serving online customers. Although serving an online customer from a store normally results in a lower net revenue due to shipping and handling costs (e.g. Bayram & Cesaret, 2021), the inventory cost are normally also lower since products do not need to be displayed on store shelves (Xu & Jackson, 2019). Little research has been conducted on how retailers should adjust to this new store operation and its impact on the retailers' performance (Mou et al., 2018). Finding an optimal rationing policy for retailers using in-store inventory for fulfilling online and offline demand is thus an important challenge. Additionally, the retailers inventory replenishment decisions are linked to this rationing policy. Therefore, the objective of this study is to identify an optimal replenishment and rationing policy for an omni-channel retailer.

Current rationing policies are not sufficient for the abovementioned retailer settings, as they typically consider a negligible lead time and the possibility to backorder. As a positive lead time and lost sales are common in retail practice (e.g. Jalilipour Alishah et al., 2015; Bayram & Cesaret, 2021), we focus in this paper on a retailer who faces a deterministic non-zero lead time, and who has to deal with lost sales in case of stock-outs. Deterministic, non-zero lead times are necessary to account for handling time and transportation efforts. Furthermore, lost sales are relevant because, when a stock-out occurs, demand cannot be backordered, as customers will often satisfy their demand elsewhere. In-store demand may only happen when products are displayed. Online customers nowadays expect immediate fulfilment of their demand with next day delivery, and they do normally not wish to backorder their demand.

Including non-zero lead times and lost sales does however increase the complexity of the decision problem, and makes an analytical approach cumbersome. Exact solution can be obtained numerically by formulating and solving the model as a Markov Decision Problem (MDP). Solving the MDP is time consuming due to the curse of dimensionality, where the amount of states for which an optimal action needs to be determined can grow exponentionally. Therefore, a faster solution would be preferred for practical implementation. In this paper, we therefore also derive a heuristic based on the structure of the optimal policy of the MDP.

We contribute to the academic literature in several ways. First, we provide an exact method for replenishment and rationing in an omni-channel bricks-and-clicks context. Second, based on our results, we are able to derive heuristics that omni-channel retailers could easily apply in practice. Third, with an extensive numerical study, we compare the heuristics with the optimal policy to find their performances.

The remainder of this paper is structured as follows. Section 2.2 presents related research on order fulfilment in an omni-channel setting and on inventory rationing. In Section 2.3, we outline the decision problem and formulate it as a MDP. In Section 2.4 , the MDP model is numerically investigated to identify the structure of the optimal policy. In Section 2.5, we derive heuristics for the ordering and rationing decision based on the structure of the optimal policies found numerically. In Section 2.6, we compare the performance of the derived heuristics in relation to the optimal policy for a wide range of instances. Section 2.7 concludes the paper and discusses future research directions.

2.2 Literature Review

Our work is related to the literature on ship-from-store strategies in omni-channel retailing and to the literature on inventory rationing. Below, we first briefly address the work on ship-from-store strategies, followed by a more comprehensive overview of the relevant inventory rationing literature.

2.2.1 Ship-from-Store Strategies

Current literature on online order fulfilment discusses the different strategies retailers can adopt to fulfil online demand. In this research, we are interested in ship-from-store strategies, a concept that uses store inventory to satisfy online demand.

One of the earliest works related to online fulfilment from store inventory is the work by Bendoly (2004). This work concluded that using store inventory for online fulfilment decreases the inventory cost for the online channel, however, satisfying in-store demand decreased due to higher stock-outs. Therefore, a trade-off between lower inventory cost and satisfying demand occurs. Further research by Bendoly et al. (2007) discussed when using store inventory is beneficial for retailers, finding that with lower percentages of online sales, store fulfilment is preferable. Ishfaq et al. (2016) also mention that using stores for online fulfilment is preferable for retailers who are aiming to adopt an online channel in a fast and inexpensive manner.

Early work on using stores for online fulfilment is mainly focused on how to allocate an online sale to different online fulfilment locations. For instance, Mahar & Wright (2009) and Mahar et al. (2009) studied a case in which the online order could be fulfilled from either a store or an online fulfilment centre. Here, the decision on where to fulfil an online order from is chosen centrally, taking into account the location of the online fulfilment centre. Bretthauer et al. (2010) and Mahar et al. (2012) extend this research by additionally taking the decision of whether a store should be included in the allocation of online orders. By not incorporating all stores for online fulfilment, their inventory can be protected from stock-outs. Similarly, Aksen & Altinkemer (2008) study store fulfilment in settings with multiple stores, deciding on which store should fulfil an online order, based on the distance to the customer and the related cost.

More recent papers focus on the order fulfilment and the operational costs. Ishfaq & Bajwa (2019) gives insight on the profitability of online fulfilment from stores when other choices of fulfilment such as vendors, distribution centres or online fulfilment centres are available. Bayram & Cesaret (2021) researches a similar setting however, includes the demand generated in-store. Difrancesco et al. (2021) specifically study the ship-from-store strategy where online orders are picked from store shelves and then shipped to the customer. Through simulation they determine the number of pickers and packers that ensures a good balance between service levels and costs.

Most literature on ship-from-store strategies is thus focused on the strategic decision on whether to adopt a ship-from-store strategy instead of adopting alternatives such as online fulfilment centres. On the operational side, the focus is often on which location should satisfy an online order. The impact of integration of channels on store operations is however not much addressed in the literature, even though an increasing number of stores are involved in online order fulfilment (Mou et al., 2018). Optimal implementation of the ship-from-store strategy is essential, as failure can lead to stock-outs, higher costs, and higher customer dissatisfaction (Difrancesco et al., 2021). Therefore, it is important for omni-channel retailers to have good inventory management and properly manage their inventory in relation to the different sales channels. In this paper, we therefore focus on inventory management related to our study which will be elaborated in the following section.

2.2.2 Inventory Rationing

Ensuring profitability for the retailer using in-store inventory for online fulfilment is an import topic in omni-channel retailing. To enable this, retailers need to develop a strategy for reserving part of their inventory specifically for online orders. Even though such a strategy is important, there is limited research on the topic. As mentioned by Jalilipour Alishah et al. (2015), Ma & Jemai (2019), and Xu & Jackson (2019) the problem setting resembles inventory rationing, a topic that has received significant attention in the literature since the seminal work by Topkis (1968). Rationing is often used to connect different types of demand with different fulfilment options, practically meaning that part of the inventory is allocated to one type of demand. These types of demand are mostly referred to as classes, which are similar to sales channels in our research. The initial rationing work by Topkis (1968) is a static strategy. Since then, research has moved from static to dynamic rationing strategies, which have been shown to be superior (Teunter & Klein Haneveld, 2008). In the remainder of this section, we therefore only focus on dynamic rationing strategies.

The dynamic rationing literature can be classified along several dimensions: replenishment policy, lead time, the number of demand classes, how shortage is dealt with, the objective of the study, and the method that is applied. The replenishment policy considers how much should be ordered and when an order should take place. Most papers consider simple policies while some consider the policy to be dynamic. A few papers do not consider replenishment in their study at all. The lead time in the studies can be either zero, a deterministic value, or a stochastic value. The shortage treatment of demand can be divided into backordering or lost sales. The number of demand classes considered is either two or a more generic N. The objective of the studies is either to reduce cost, increase profit, or improve service level. The method the authors applied can be differentiated in MDP, simulation, or mathematical analysis. Table 2.1 presents a chronologically ordered overview of the literature on dynamic rationing, including the classification on the six dimensions.

Replenishment policy and lead time

From Table 2.1, it is observed that the first papers investigating dynamic rationing only considered an (s, Q) replenishment policy, which is the optimal replenishment strategy for static rationing according to Ha (1997). Not all papers consider a replenishment policy. Some papers only investigate a single period with a fixed initial inventory level, while other consider a manufacturing system in which some production planning decisions are considered instead (Carr & Duenyas, 2000; Turgay et al., 2015).

Wang & Tang (2014) concluded that, for periodic review, a base-stock replenishment policy is the optimal ordering policy in situations with zero lead time. Chew et al. (2013) were the first to study a dynamic replenishment policy and concluded that the optimal dynamic replenishment policy resembles a base-stock replenishment policy, which is confirmed in later studies (e.g., Bao et al., 2018; Xu & Jackson, 2019). However, optimal replenishment policies in situations with non-zero lead times could not be derived due to the complexity of the problem.

The replenishment policy is relevant for the rationing decision, as the rationing decision is based on current and future inventory levels. Thus outstanding replenishment orders influence the rationing decision. It can be concluded that the optimal dynamic replenishment policy for situations with deterministic lead times has yet to be addressed in the literature.

Table 2.1: Literature review of dynamic inventory rationing.

Paper	Replenishment: Order policy	Replenishment: Lead time*	Demand: Back-ordering	Demand: Lost sales	Demand: Classes	Optimisation focus: Cost	Optimisation focus: Profit	Optimisation focus: Service level	Method: MDP	Method: Simulation	Method: Mathematical analysis
Kaplan (1969)			×		2	×					×
Haynsworth & Price (1989)	$(s,Q)^\dagger$	D	×		2			×		×	
Ha (1997)			×		2	×			×		
Carr & Duenyas (2000)				×	2		×		×		
Melchiors et al. (2000)	$(s,Q)^\dagger$	D		×	2	×					×
Melchiors (2001)	$(s,Q)^\dagger$	S		×	N	×			×		
Deshpande et al. (2003)	(s,Q)	D	×		2	×					×
Frank et al. (2003)	(s,S)	Z		×	2	×					×
Melchiors (2003)	$(s,Q)^\dagger$	D		×	N	×			×		
Teunter & Klein Haneveld (2008)	(s,Q)	Z	×		2	×					×
Gayon et al. (2009)		S	×		N	×					×
Benjaafar et al. (2010)			×	×	2	×			×		
Fadıloğlu & Bulut (2010)	(s,Q)	D	×	×	N	×				×	
Zhao & Lian (2011)	(s,Q)	S	×		2	×					×
Chen & Bell (2012)	(s,Q)	S		×	2	×					×
Hung et al. (2012)	(R,S)	D	×		N	×					×
Chew et al. (2013)	Dynamic	Z & D	×		N	×			×		
Hung & Hsiao (2013)	$(s,S)^\dagger$	S		×	N			×		×	
Wang et al. (2013a)	(s,Q)	D	×		2	×		×			×
Wang et al. (2013b)		S	×		2	×			×		
Wang & Tang (2014)	(R,S)	Z	×	×	N	×			×		
Liu & Zhang (2015)			×		2	×			×		
Liu et al. (2015)	(R,S)	Z	×		N	×					×
Turgay et al. (2015)				×	N		×				×
Alfieri et al. (2017)	$(R,S)^\dagger$	D	×		2	×				×	
Bao et al. (2018)	Dynamic	Z	×		N	×					×
Xu & Jackson (2019)	Dynamic	Z	×		2		×				×
This study	Dynamic	D		×	2		×		×		

*D = Deterministic, S= Stochastic, Z= Zero † Fixed parameters for replenishment policy

Demand

Wang & Tang (2014) were the first to compare the backordering and lost-sales setting. From the comparison, it was found that in a backordering setting the rationing level increases occasionally to ensure inventory for future demand. With lost sales, the rationing level decreases when coming closer to replenishment. In general, considering lost sales increases the complexity of the system, as the inventory level does not change with unmet demand (Frank et al., 2003; Fadıloğlu & Bulut, 2010; Chen & Bell, 2012; Hung & Hsiao, 2013).

Melchiors (2001) was the first to propose a model with N demand classes, although the numerical results are only presented for 2 demand classes. Most papers propose a model for N demand classes, but are not able to derive exact solutions if the demand classes exceeds 2 due to the complexity (Melchiors, 2001; Chew et al., 2013). Bao et al. (2018) is able to derive an exact solution for N demand classes but mentions introducing non-zero lead time makes finding the optimal policy much more complicated.

Optimisation focus

The rationing decision is often based on minimising costs. This is also related to how shortages are considered: if demand is always satisfied (through backordering), the rationing decision only influences cost and not profit. Studies that consider lost sales and base their rationing on minimising cost, consider a shortage cost for the lost sales. By applying different shortage costs, they are able to achieve different service levels for the demand classes (Chew et al., 2013; Liu et al., 2015).

Carr & Duenyas (2000) and Turgay et al. (2015) were the first to base their rationing decisions on profit maximisation for a make-to-stock production system, concluding that the structure of the optimal policy is complex, and therefore exploring the solution with a simpler policy. Xu & Jackson (2019) used profit to seek a balance between fulfilling online demand and the related high fulfilment cost. In general, using profit as optimisation focus increases the complexity of the structure of the optimal policy. However, it better captures the omni-channel fulfilment strategies as it allows the balancing between satisfying in-store and online demand.

Method

In most papers, dynamic programming is used to find the optimal solution and find the structure of the rationing policies. However, as mentioned previously, the structure of the system can become complex for certain system characteristics. As the problem easily become multi-dimensional, the curse of dimensionality results in complex and not insightful formulas (Teunter & Klein Haneveld, 2008). Therefore, research often resorts to finding near-optimal policies for the problem (Liu & Zhang, 2015).

By applying MDP to solve the dynamic rationing model, an exact solution can be derived. Most literature using MDP to find the optimal solution also uses the structure of the solution to derive simple heuristics (e.g. Wang et al., 2013a; Chew et al., 2013) as MDP can require high computational time.

2.2.3 Knowledge Gap

The dynamic rationing problem resulting from the ship-from-store context is characterised by deterministic lead times and lost sales. Because online demand cannot be backlogged as the customer requires same day shipment, shortages lead to lost sales. As the in-store and online channel have different cost and benefits, only the profit is able to encompass the trade-off between the different channels.

In the literature review we can see that the paper by Chew et al. (2013) is closely related as it applies a dynamic order policy with deterministic lead time. They do however not include lost sales, but mentions its importance for further research. Melchiors (2001, 2003) and Wang & Tang (2014) are also related as they investigate the lost sales problem using MDP. Wang & Tang (2014) specifically mention the investigating of different order policies and the inclusion of lead time as future research. Although Melchiors (2001, 2003) does add lead time in their research, they limit themselves by fixing the order policy. Finally, Xu & Jackson (2019) study an omni-channel retailer in a similar setting, but acknowledge that their assumption of no lead time limits their findings.

Based on these findings, we conclude that no lost sales model exits that integrates the ordering and rationing decision for a profit maximising retailer who serves both an online and an offline demand from a stock point with non-zero replenishment lead time. We contribute to the existing literature by deriving heuristics and compare them with the exact solution obtained from the MDP. In the next section we present the problem and a model to derive an exact solution.

2.3 Markov Decision Process

2.3.1 Problem Definition

The problem that is studied is determining an optimal rationing and replenishment policy. A typical setting for a retailer is that they can order new products every week but can ration their products on a daily basis. Managing the in-store inventory thus happens quite regularly, while ordering and replenishment are more dependent on fixed delivery schedules. The rationing decision at the start of every day is motivated by practices in which the online sales will be packed and shipped at the end of a day. Which is very common to happen after some cut-off time for online orders. The rationing decision is needed to ensure that enough products are left to fulfil the online orders placed during the day. These two decision problems can be formulated as a hierarchical decision problem. The problem has two levels, where at level I the replenishment decision is made and at level II the rationing decision is made. The time between two ordering decisions (at level I) is called a period. As the rationing decision at level II is taken more frequent, level II is split into $R = 7$ sub-periods.

The objective of the problem is to maximise the profit resulting from sales revenues on the one hand and holding costs, shipment costs, and procurement costs on the other hand. We differentiate two types of customers: those who visit the offline channel (the physical store) and those who use the online channel. It is assumed that there is no channel substitution and that the cost of the online and offline channels differ. Shipment costs only applies to online sales. Nowadays, customers expect free delivery thus the costs are carried by the retailer. The sales price in the two channels are the same in an omni-channel setting.

In Figure 2.1 the problem is presented with its two levels, where at level I the retailer makes the replenishment decision which is based on the actual inventory level. At level II, the rationing decision is set based on the actual inventory level and the outstanding order (which was set at level I). Every week (M) the replenishment decision is made at the start of the first day, which will be delivered after a fixed lead time L. The replenishment occurs at the start of the delivery day L. At the start of every day, the inventory is rationed. The rationing decision a_t sets how much products are made available for the offline channel on day t, with the remainder made available for the online channel. We do not assume that throughout the day the retailer checks for ex-

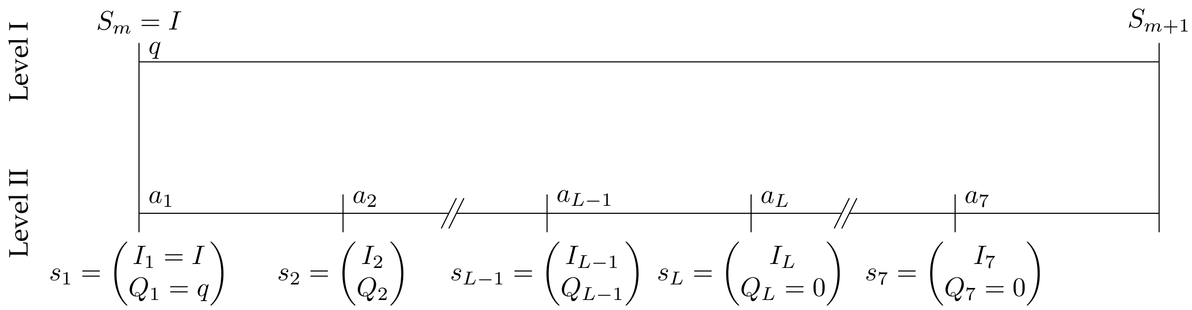


Figure 2.1: Visualisation of the actions for the hierarchical problem.

cess stock in one channel if the other channel has a stockout. A product sold through the online channel is not directly removed from stock (as shipment occurs at the end of the day), thus the retailer needs to be certain the product is not already sold. Additionally, online customers might have their product in their online basket, thinking the product is available while the retailer might be using their product to fulfil in-store demand. Furthermore, it is difficult to capture lost demand, as customers who face empty shelves often walk out of the store. Hence to avoid lost sales in the offline channel by taking a product from the online channel requires keeping track of the in-store inventory level. But as the retailer often has many products, continuously replenishing the store from the back storage is cumbersome and might also influence the shopping experience of customers. For the replenishment decision we assume the retailer orders every $R = 7$ days and faces a lead time (L) of at most 7 days. After $R = 7$ days, the process repeats with the inventory level at the start of week $m + 1$, equal to the closing inventory of week m.

2.3.2 MDP model

The problem described above can be formulated as a MDP, where the demand in the channels introduces uncertainty in the MDP. With the ordering and rationing decision the retailer can control its inventory levels in both channels. As the MDP consists of two different decisions on different time periods, it can be integrated in a Hierarchical Markov Process as described in Kristensen (1988).

States and State Spaces

The state at level I of the MDP consists of the inventory level I at week m and is given as $S_m = I$. It is assumed that the retailer will not have more products in store than the maximum demand until the next replenishment. As holding cost are positive and no fixed ordering cost applies, there is no motivation for the retailer to order more than the maximum demand as excessive stock would negatively influences the profit. Therefore, the state space at level I is $I \in \mathcal{I}$ with $\mathcal{I} = \{0, 1, ..., (7 + L) \cdot D\}$, in which the maximum demand of a day is $D = \sum_i D_i$. D_i indicates the maximum possible demand of the individual channels with $i \in \{1 = \text{offline}, 2 = \text{online}\}$. The demand in both channels during each sub-period is modelled by Poisson distributions, $P_i(d_i)$ is the probability that the demand of channel i is d_i where $d_i \in \{0, 1..., D_i\}$.

The state at level II consist of the inventory level and the outstanding replenishment quantity and is given as $s = (I_t, Q_t)$. The state space of the inventory at level II is the same as at level I, $I_t \in \mathcal{I}$. The state space of the replenishment quantity depends on expected demand and the current inventory level, $Q_t \in \mathcal{Q}_t(I_t)$ with $\mathcal{Q}_t(I_t) = \{0, 1, ..., 7 \cdot D - I_t\}$. The upper bound of the state space is set under the assumption that the retailer will not replenish more products than total maximum expected demand per 7 days minus the current inventory.

Actions and Action Spaces

At level I of the MDP, a replenishment quantity q is set at the beginning of the period. The action space is thus equal to the state space of the replenishment quantity at level II thus $q \in \mathcal{Q}_t(I_t) = \{0, 1, ..., 7 \cdot D - I_t\}$.

At level II of the MDP, the rationing across the inventory a_t is set. It indicates how many products of the total inventory is withheld from the online channel and placed in the offline channel. The action space of a_t is dependent on the inventory I_t at sub-period t since the rationing quantity is clearly limited to the current inventory. Thus, a_t can be defined as $a_t \in \mathcal{A}_t(I_t)$ with $\mathcal{A}_t(I_t) = \{0, 1, ..., I_t\}$. We assume that the rationing decision is made at the beginning of the sub-period and will not be revised during the sub-period.

Transitions

At level I of the MDP, the state only transitions at the end of the week, while at level II the states transition every day. At level II of the MDP, we model the transition from state $s_t = (I_t, Q_t)$ to $s_{t+1} = (I_{t+1}, Q_{t+1})$. The transition of I_t to I_{t+1} depends on the current inventory, the rationing decision, demand of the individual channels, and the inventory replenishment:

$$I_{t+1} = (a_t - d_1)^+ + (I_t - a_t - d_2)^+ + \delta(L = t) \cdot Q_t \tag{2.1}$$

Where $x^+ = \max(x, 0)$ and $\delta(x)$ denotes the Kronecker delta, which returns the value 1 if $x =$ True, otherwise 0. Equation (2.1) refers to the inventory being rationed across the two channels and the replenishment, with a_t and $I_t - a_t$ being the inventory level of the individual channels at weekday t from which the fulfiled demand of the individual channels is subtracted from. The replenishment takes place at $t = L$, which is at the beginning of day L.

The transition of Q_t to Q_{t+1} depends on the weekday. At the first weekday the state Q_1 takes the value of replenishment quantity q. Q_{t+1} remains Q_t until the replenishment is added to the stock at time L after which it is set to zero:

$$Q_{t+1} = \begin{cases} Q_t & \text{if } 1 < t \leq L \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

At level I of the MDP, the next state is noted by S_{m+1} which equals the closing inventory of the week m, which is obtained from the last state transition on level II of the MDP:

$$S_{m+1} = (a_7 - d_1)^+ + (I_7 - a_7 - d_2)^+ + \delta(L = 7) \cdot Q_7 \tag{2.3}$$

Expected immediate reward

The objective of the MDP is to maximise the profit (over an infinite horizon) which consists of revenue generated from selling products, and the cost of shipment, holding, and replenishment. At level I the expected immediate reward of the MDP only consists of the replenishment cost:

$$\mathbb{E}C_{\mathrm{I}}(q) = -c_p \cdot q \tag{2.4}$$

The replenishment cost is based on a variable replenishment cost c and the replenishment quantity q.

For level II of the MDP the expected immediate reward depends on the revenue from selling products, cost of shipment and holding cost of the individual channels. The expected profit depends on the action a_t taken in state $s_t = (I_t, Q_t)$:

$$\mathbb{E}C_{\text{II}}(s_t, a_t) = \begin{cases} p\left(\sum_{d_1<a_t} d_1 \cdot P_1(d_1) + \sum_{d_1 \geq a_t} a_t \cdot P_1(d_1)\right) & \text{(2.5a)} \\ +\,(p-c_u)\left(\sum_{d_2<I_t-a_t} d_2 \cdot P_2(d_2) + \sum_{d_2 \geq I_t-a_t} (I_t-a_t) \cdot P_2(d_2)\right) & \text{(2.5b)} \\ -\,(c_{h1} \cdot a_t + c_{h2} \cdot (I_t - a_t)) & \text{(2.5c)} \end{cases}$$

The first term (2.5a) is the revenue from the offline channel for price p and the second term (2.5b) the revenue from the online channel. The same sales price is applied but when satisfying an online demand from the store a unit shipment cost u is incurred. The quantity sold through one channel depends on the demand and rationing. Shortage cost is not included into this model, if shortage occurs the retailer faces a lost sales causing in a penalty by losing profit margins. Next to shipment costs, the model accounts for the holding cost (2.5c).

2.3.3 Value iteration

The aim of the MDP is to maximise the long-term weekly expected profit. With value iteration, the problem is solved iteratively backwards, where one iteration relates to the a single sub period, e.g. a day. Hence a sequence of $R = 7$ iterations relate to one week. To ease the explanation of the algorithm, the two levels of the MDP are integrated into one level.

Let n be the iteration counter. The maximum expected profit over n consecutive days when starting in state s_t is defined as $v_n(s_t)$. The long run weekly profit is thus $g = \lim_{n\to\infty} (v_n - v_{n-7})$, which does not depend on the initial state s, and is thus the same for all s (the value of g is called the gain of the underlying Markov chain). To determine the value of g by value iteration one starts setting $v_0(s) = 0$ for all states s. Next, one computes for all $s : v_1(s_t) = \max_{a_t \in \mathcal{A}_t(I_t)} \{\mathbb{E}C_{\text{II}}(s_t, a_t)\}$, and continues by computing v_2, v_3, etc. using the so called recursive Bellman equations in (2.6) and (2.7). The day t can be calculated from the iteration n as follow: $t = (7-(n-1) \mod 7)$. The distinction of day t is relevant to keep track of whether next to the rationing decision an order decision should be taken or an outstanding order has arrived.

For $t = 1$, the Bellman equation incorporates ordering and rationing decision and the respective costs and revenues to maximise the expected profit:

$$\begin{aligned} v_n(s_1) = \max_{q \in \mathcal{Q}_1(I_1)} \Bigg\{ & \mathbb{E}C_{\mathrm{I}}(s_t, q) \\ & + \max_{a_1 \in \mathcal{A}_1(I_1)} \Bigg\{ \mathbb{E}C_{\mathrm{II}}(s_1, a_1) + \sum_{d_1=0}^{D_1} \sum_{d_2=0}^{D_2} P_1(d_1) P_2(d_2) v_{n-1}(s_2) \Bigg\} \Bigg\} \end{aligned} \tag{2.6}$$

For $t = 2, 3, ..., 7$ the Bellman equation has to consider the rationing decisions:

$$v_n(s_t) = \max_{a_t \in \mathcal{A}_t(I_t)} \Bigg\{ \mathbb{E}C_{\mathrm{II}}(s_t, a_t) + \sum_{d_1=0}^{D_1} \sum_{d_2=0}^{D_2} P_1(d_1) P_2(d_2) v_{n-1}(s_{t+1}) \Bigg\} \tag{2.7}$$

As $n \to \infty$, the span of the average weekly expected profit decreases, thus $\|v_n - v_{n-7}\|$ converges to 0, which implies that the difference between the largest and the smallest element of $v_n - v_{n-7}$ becomes zero, and all elements equal the maximum expected long run weekly profit. If $\|v_n - v_{n-7}\|$ is smaller than ε the value iteration stops, we specify $\varepsilon = 0.001$. Algorithm 1 formalises the algorithm and shows how value iteration can be implemented. In particular it demonstrates that the two maximisation actions in equation (2.6), can be implemented in serial to reduce the computation complexity of value iteration. The algorithm applies backward induction. The first for loop deals with the rationing decisions at level II. Next the ordering decision is optimised and the procurement costs are added to v_n. As the average weekly expected profit converges to an vector consisting of equal values, the value iteration is stopped.

Algorithm 1: Value iteration of the MDP.

initialisation: $n = 0$; $v_0 = 0$;

repeat

  for $t = 7 : -1 : 1$ **do**

    $n = n + 1$

    for $s_t = (I_t, Q_t) \in \{(I_t, Q_t)|\ I_t \in \mathcal{I}, Q_t \in \mathcal{Q}_t(I_t)\}$ **do**

      $v_n\left(s_t\right) = \max\limits_{a \in A_t(I_t)} \left\{ \mathbb{E}C_{\mathrm{II}}(s_t, a) + \sum\limits_{d_1=0}^{D_1} \sum\limits_{d_2=0}^{D_2} P_1\left(d_1\right) P_2\left(d_2\right) v_{n-1}\left(s_{t+1}\right) \right\}$ where:

      s_{t+1} is from equation (2.1) and (2.2)

    end

  end

  for $I_1 \in \mathcal{I}$ **do**

    $w(I_1) = \max\limits_{q \in \mathcal{Q}(I_t)} \left\{ \mathbb{E}C_{\mathrm{I}}(q) + v_n\left(s_1\right) \right\}$ where: $s_1 = (I_1, q)$

  end

  for $I \in \mathcal{I}$ **do**

    $v_n(I, 0) = w(I)$

  end

until $\|v_n - v_{n-7}\| < \epsilon$;

The (nearly) optimal strategy for the two levels of the MDP can be obtained from the results of the value matrices. First the optimal replenishment policy $\pi^{\mathrm{I}}\left(s_1\right)$ at $t = 1$ for level I is found by equation (2.8):

$$\begin{aligned}\pi^{\mathrm{I}}(s_1) = \arg\max_{q \in \mathcal{Q}_1(I_1)} \Bigg\{ & \mathbb{E}C_{\mathrm{I}}\left(s_t, q\right) \\ & + \max_{a_1 \in \mathcal{A}_1(I_1)} \left\{ \mathbb{E}C_{\mathrm{II}}\left(s_1, a_1\right) + \sum_{d_1=0}^{D_1} \sum_{d_2=0}^{D_2} P_1\left(d_1\right) P_2\left(d_2\right) v_{n-1}\left(s_2\right) \right\} \Bigg\}\end{aligned} \tag{2.8}$$

Second, the optimal rationing decision $\pi_t{}^{\mathrm{II}}(s_t)$ for all states in $t = 1, 2, ..., 7$ for level II is found by equation (2.9):

$$\pi_t{}^{\mathrm{II}}\left(s_t\right) = \arg\max_{a_t \in \mathcal{A}_t(I_t)} \left\{ \mathbb{E}C_{\mathrm{II}}\left(s_t, a_t\right) + \sum_{d_1=0}^{D_1} \sum_{d_2=0}^{D_2} P_1\left(d_1\right) P_2\left(d_2\right) v_{n-1}\left(s_{t+1}\right) \right\} \tag{2.9}$$

2.4 MDP Computational Complexity and Results

2.4.1 Data and design of experiments

The optimal policy is investigated for a range of instances, consisting of a base test case and various alterations. The data set used is based on recent literature on similar omnichannel retail environments (Jalilipour Alishah et al., 2015; Li et al., 2015; Ovezmyradov & Kurata, 2019; Xu & Jackson, 2019; Bayram & Cesaret, 2021). We alter different parameters to investigate the performance of the MDP. In Table 2.2, the parameter values are found for the review periods, lead time, mean daily demand, cost of product, fulfilment cost, and holding cost of the channels are given. Without loss of generality, we set $p = 100$ and $c_{h2} = 0.5$, and vary the other cost parameters in relation to these fixed parameters. The different instances are all created by varying specific subsets of parameters.

One typical instance is used as a base test case. The demand is independently Poisson distributed with a mean daily demand of $\mu_1 = 6$ and $\mu_2 = 2$, with a maximum demand of $D_1 = 12$ and $D_2 = 6$. In line with related papers, such as Jalilipour Alishah et al. (2015); Xu & Jackson (2019); Bayram & Cesaret (2021), we assume no substitution between offline and online demand. Replenishment orders are placed on Monday ($t = 1$) and delivered on Wednesday ($t = 3$), where both events occur at the beginning of the day. The cost of the product $c_p = 30$, the handling cost incurred for satisfying an online order with store inventory is $c_u = 5$. The holding cost of the offline is $c_{h1} = 1$. In the remainder of this paper, if an instance does not specify a certain parameter value, it will be equal to their base test case setting.

Table 2.2: Parameter values, where the base case is given in bold.

Parameter	Symbol	Values
Offline demand	μ_1	$2, 4, \mathbf{6}$
Online demand	μ_2	$\mathbf{2}, 4, 6$
Review period	R	$2, 3, 4, 5, 6, \mathbf{7}$
Lead time	L	$1, \mathbf{2}, 3, 4, 5, 6, 7$
Price	p	$\mathbf{100}$
Procurement cost	c_p	$20, \mathbf{30}, 40$
Shipping cost	c_u	$0, \mathbf{5}, 20$
Offline holding cost	c_{h1}	$\mathbf{1}, 2, 3$
Online holding cost online	c_{h2}	$\mathbf{0.5}$

2.4.2 Computational implementation

The number of transitions from one state to the other depends on the action space and demand distribution. The action space is already limited by inventory and demand level. As is common in numerical analysis of MDPs, we truncate the demand distribution to cover 99% of the cumulative distribution function to limit the state space. The demand distribution is reshaped as a right-truncated Poisson distribution as described in Cohen (1954).

Additionally, to ensure that enough memory is available to compute all the expected reward matrices, not all matrices are stored. Matrices at level I older than $n - 1$ are deleted. At level II, matrices older than $t + 1$ are not stored. However, to find the optimal rationing policy, matrices older than $t + 1$ are needed to derive the optimal policy. This is solved by storing the action a_t in $\pi^{\text{II}}(s_t)$ after the second Bellman equation. In the last, iteration the optimal rationing strategy is approximated by the action $\pi^{\text{II}}(s_t)$.

Since the transition probabilities require high computation time, they are calculated beforehand and stored in a matrix. This ensures that all transition probabilities are only calculated once. As the amount of transition probabilities is only $4.3 \cdot 10^6$, it is possible to store them.

For solving the MDP, the policy needs to be calculated for a large number of states. The results were obtained by implementing the MDP in Python version 3.7.2. The model was run on a Personal Computer with Intel Xeon W-2133 CPU @ 3.60 GHz and 16GB of RAM. Table 2.3 shows for all instances the total number of states, the CPU time in seconds, and RAM usage in megabyte (MB) for the MDP. The instances with different cost structures are not presented as they do not increase the complexity of the model.

From Table 2.3 it is observed that an increase in lead time or review period increases the computational complexity of the MDP. A longer review period increases the inventory and ordering state space. A longer lead time means that the ordering state is included in more states, increasing the total amount of states. The total amount of states grows linearly with the lead time, but exponentially with the review period. The time to solve the MDP grows with the number of states; as the problem becomes bigger, more RAM is needed to solve an instance. A higher average daily demand also increases the complexity of the MDP, as the maximum demand increases with a higher average

Table 2.3: Computational complexity of the MDP for different instances.

Instances	States	CPU time (s)	RAM (MB)
$L = 1$	127890	180	52.56
$L = 2$*	143766	448	74.14
$L = 3$	159642	810	101.00
$L = 4$	175518	1570	133.72
$L = 5$	191394	2710	172.89
$L = 6$	207270	3780	219.10
$L = 7$	223146	5110	270.41
$R = 2$	5256	33	6.33
$R = 3$	14742	76	12.75
$R = 4$	31392	144	22.00
$R = 5$	57150	204	34.91
$R = 6$	93960	321	52.09
$R = 3, L = 3$	17658	163	21.33
$R = 4, L = 4$	41760	505	50.52
$R = 5, L = 5$	81450	1430	98.61
$R = 6, L = 6$	140616	2970	170.33
$\mu_1 = 2, \mu_2 = 2$	64092	173	25.43
$\mu_1 = 4, \mu_2 = 4$	143766	478	65.01
$\mu_1 = 2, \mu_2 = 6$	143766	478	74.14

*Base test case with $R = 7, L = 2, \mu_1 = 6, \mu_2 = 3$,
$p = 100, c_p = 30, c_u = 5, c_{h1} = 1, c_{h2} = 0.5$

demand. As the required number of states for the inventory and ordering decisions is related to the maximum demand, the total amount of states increases with a higher average daily demand.

Table 2.3 shows that solving the MDP can take more than one hour, which is impractical for retailers. To avoid long computation time heuristics are preferred. By analysing the structure of the optimal policy, we can derive effective rules that are able to encompass the complexity of the MDP.

2.4.3 Structure of the optimal policy

We numerically investigate the optimal policy and derive general rules from the structure of the policy so that the heuristics are applicable for all instances presented in Table 2.2. The structure of the optimal replenishment policy and rationing policy is identified for the base test case.

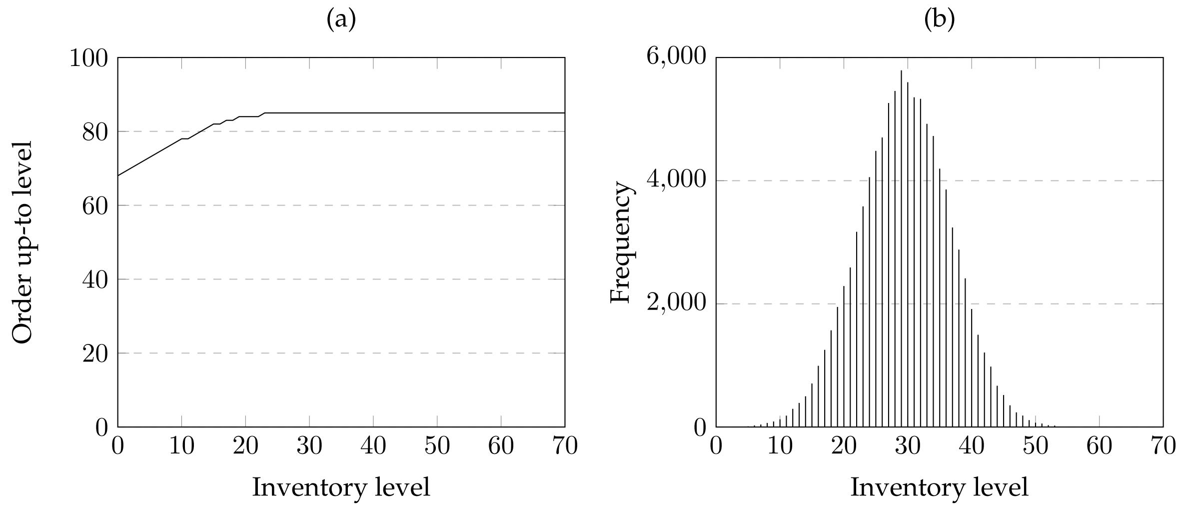


Figure 2.2: Order-up-to level (a) and inventory frequency (b) for different inventory levels at the beginning of day 1.

Replenishment Policy

Figure 2.2(a) shows the order-up-to level at the day when orders are placed for different inventory levels. The order-up-to level is given by adding the ordered quantity to the inventory level. From the figure it can be concluded that at lower inventory levels the optimal replenishment policy resembles an (R, Q) replenishment policy, as the replenishment quantity remains (almost) the same. At higher inventory levels, it resembles a base stock policy as observed by the maximum order-up-to level at inventory levels above 22.

In order to get better insight in which order-up-to levels occur more frequently the optimal policy is simulated for 100,000 weeks. From the simulation the frequency of different inventory levels can be found, as seen in Figure 2.2(b). It is seen that both replenishment policies occur with the base stock policy being more common. As the average demand is $\mu_1 + \mu_2 = 8$, and the inventory being below 8 is 0.2%, the chance of a lost sale is small.

By using an approach similar to Haijema et al. (2007), a frequency table is used to find the structure of the optimal policy for different inventory and order-up-to levels. Table 2.4 shows the simulation results of the frequency of each order-up-to level with inventory level. The total frequency of each order-up-to level is presented in the last column. From this table, the optimal order quantities can also be derived. For instance, when looking at the columns for inventory level 17 and 18, we can see that in 100,000

Table 2.4: Frequency table of order-up-to level with inventory level.

Inventory	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	(23,24,...,70)	Total
Up-to 85																								83733	83733
84																				1950	2289	2591	3170		10000
83																		1523	1570						2823
82																707	994								1701
81															499										499
80														391											391
79													295												295
78											125	185													310
77										90															90
76									66																66
75								41																	41
74							26																		26
73						11																			11
72					8																				8
71				2																					2
70			1																						1
69		3																							3
68	0																								0
..																									..
0																									0

simulations, there were 1,523 occasions in which the retailer had an inventory level of 17 upon ordering, and 1,570 occasions in which the inventory level was 18. For both these cases, the optimal order-up-to level is 83 (represented in the row).

From Table 2.4 it is observed that for inventory levels 0 to 10 no fixed order-up-to level but a fixed order quantity of 68 applies. This indicates that the optimal policy expects and anticipates that the available inventory will be sold during the procurement lead time. At higher inventory levels there is a chance that leftovers occur upon replenishment, thus lower quantities are ordered. When plenty of inventory is left, an uncertain amount is carried over to the next week. For stock levels higher than 22 a base stock policy applies, which occurs around 85% of the time. At most 17 products are reserved for demand during lead time, as the base stock is 85 and the optimal inventory level is 68.

The behaviour of the optimal replenishment policy can be explained by the lead time. Current orders that are placed will be replenished after the lead time has exceeded. The optimal inventory level when orders are replenished is 68, thus the replenishment policy tries to achieve this. At low inventory levels it is expected that the stock is depleted at $t = L$, thus 68 products are ordered, following an (R, Q) order policy. At higher inventory levels, a certain amount of left-over products is expected at $t = L$, and to compensate for these left-overs they are subtracted from the order quantity, resembling an (R, S) order policy.

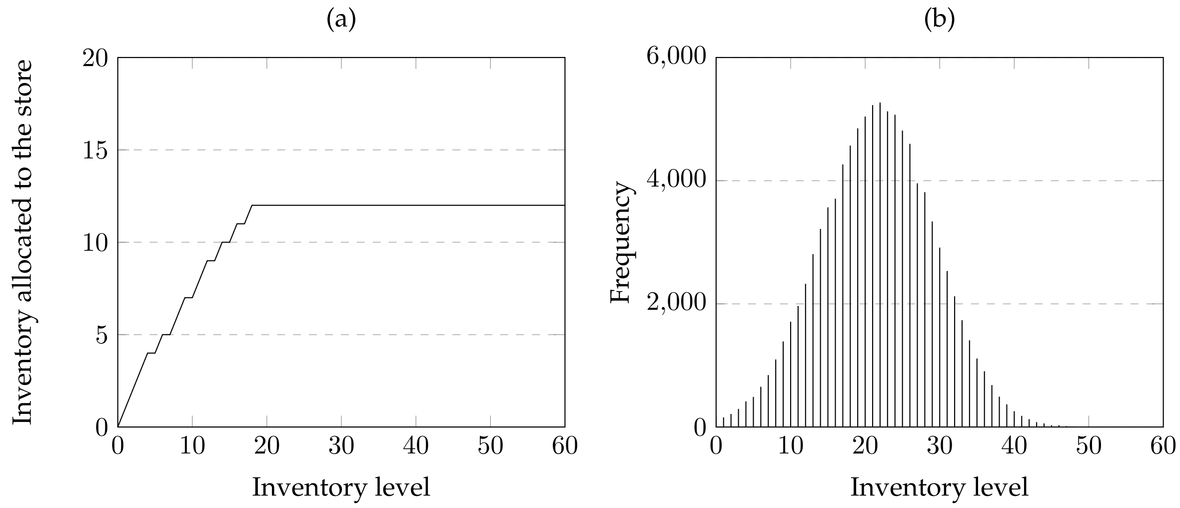


Figure 2.3: Rationing policy (a) and inventory distribution (b) of different inventory levels at the beginning of day 2 if the order quantity is 62.

Rationing Policy

Figure 2.3(a) displays the optimal rationing policy at day 2, which is the day before replenishment. Day 2 is chosen for this analysis, as the day before replenishment inventory levels are lowest, thus rationing is most important. The order quantity 62 is chosen, as it is the most frequent order quantity. At high inventory levels, the amount of in-store inventory is a fixed amount as additional in-store inventory would not be worth the additional costs. This base stock level can also be proven analytically, which can be seen in Appendix 2.A. At low inventory levels, there is a choice between whether to store products in-store or offline. This trade-off is based on different drivers such as expected future sales and holding cost. It is seen that the trade-off between storing a product in-store or online is almost a fixed ratio.

From Figure 2.3(b) it is seen that there is a risk of shortage, as the expected average daily demand is 8 and inventory levels below 8 occur in a significant number of instances. For 95.7% of the weeks, the inventory is able to satisfy the average daily demand at day 2. For 4.3% of the weeks, the inventory is however below 8. Comparing Figure 2.2(b) to Figure 2.3(b), the average inventory level shifts from 30 to 22. This shift is expected for the inventory levels of two consecutive days if the average daily demand is 8.

2.5 Heuristics

Solving the MDP for the base case takes around eight minutes for a single product and can take over an hour for instances with long lead times. Heuristics to reduce computational time are then preferred, especially when a retailer manages multiple products. Therefore, we develop heuristics for the ordering and rationing decision that approximate the structure of the optimal policy identified above.

From Figure 2.2 and 2.3, it is clear that the ordering and rationing decisions turn out to have a threshold. When the inventory level is below the threshold, the optimal ordering decision can be approximated by, a fixed order quantity. At high inventory levels, a base-stock policy seems to fit. For the rationing decision, at low inventory levels a trade-off exists between the channels which is driven by expected future sales and holding costs. At high inventory levels, we already identified an optimal strategy which consists of a fixed amount of products stored in-store (found in Appendix 2.A). In the following, these results and insights are used to develop ordering and rationing heuristics.

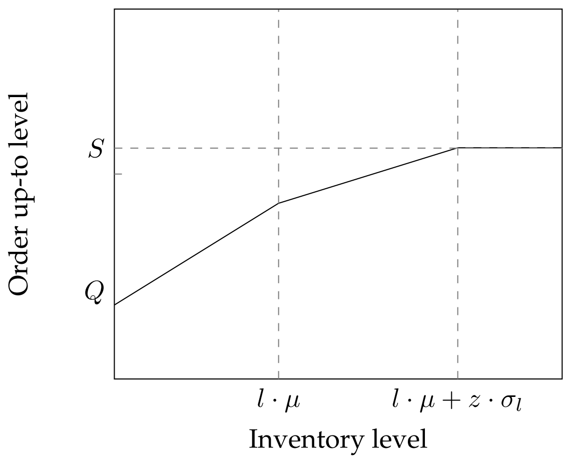
2.5.1 Ordering Heuristic

The heuristic for ordering is based on having a fixed order quantity at low inventory levels and a base-stock policy at high inventory levels, as it was found in Figure 2.2 to be optimal. In between the low and high inventory levels the optimal ordering policy is a combination of a fixed order quantity and base-stock policy.

Figure 2.4(a) illustrates the structure found for the optimal ordering policy. In the figure, two inventory level points are presented which are used to derive a heuristic for ordering. The first inventory level point indicates that if the inventory is below the level it is assumed there is no inventory when replenishment occurs. This inventory level point is found from the average sales procurement lead time ($L \cdot \sum_i \mu_i$), as this is the amount of products that likely will be sold before replenishment occurs.

The second inventory level point indicates the point in which it is certain there is left-over inventory at replenishment. This is given if the inventory level exceeds the average sales with safety stock procurement lead time ($L \cdot \sum_i \mu_i + z \cdot \sigma_L$), it is unlikely the retailer will sell more products during lead time before replenishment occurs. The safety stock z follows the critical fractile ratio $\left(\frac{p-c_p}{p-c_p+R \cdot c_{h2}}\right)$. As the safety stock is often

(a)



(b)

Algorithm 2: Heuristic for ordering.

$z = F^{-1}\left((p - c_p) \,/\, (p - c_p + R \cdot h_2)\right)$
$\mu = \sum_i \mu_i$
$Q = R \cdot \mu + z \cdot \sigma_R$
$S = (R + L) \cdot \mu + z \cdot \sigma_{R+L}$
if $I < L \cdot \mu$ **then**
  $q = Q$
else if $L \cdot \mu \leq I \leq L \cdot \mu + z \cdot \sigma_L$ **then**
  $w = (I - L \cdot \mu)/(z \cdot \sigma_L)$
  $q = \text{round}\left((1 - w) \cdot Q + w \cdot (S - I)\right)$
else
  $q = S - I$
end

Figure 2.4: Visualisation of the parameters of the ordering heuristic (a) with the algorithm (b).

a formula consisting of the lost sale cost and holding cost, it is assumed in this heuristic that the lost sale cost is equal to the profit of selling a product in the offline channel. Reason is that the optimal policy is to sell the product through the offline channel. The holding cost is assumed to be from the online channel, since excess stock is preferred to be stored in the online channel. F^{-1} denotes the inverse of the Normal distribution function with mean zero and variance of one. σ_i is the standard deviation of the total demand during period i.

The region between the two inventory level points indicate there is uncertainty if there will be inventory left-over at replenishment or not. If the inventory level is closer to the average sales procurement lead time it is more likely that at replenishment the inventory level is zero. However, if the inventory is closer to the average sales with safety stock procurement lead time there is a higher change of left-overs at replenishment.

The region which assumes that no inventory is left at replenishment follows an fixed order quantity policy, where Q is the order quantity as seen in Figure 2.4(a). The order quantity is derived from the average sales with safety stock during the review period and is calculated as follow: $R \cdot \sum_i \mu_i + z \cdot \sigma_R$.

If the inventory level exceeds the average sales with safety stock procurement lead time, it is certain that there is left-over inventory at replenishment thus a base-stock policy is used. The order-up-to level S is derived from the average sales during the review period and lead time. The order-up-to level uses the same formula as the fixed order quantity however, takes the average sales and safety stock over the review period plus lead time: $(R + L) \cdot \sum_i \mu_i + z \cdot \sigma_{R+L}$.

In between the fixed order quantity and base-stock policy, the optimal policy is a combination of the two. Therefore, we linearly interpolate the two policies in the region where the inventory level is between the two inventory level points. By using a weight ratio (w) we decide whether the order policy should follow more a fixed order quantity or base-stock policy. The weight ratio is a linear interpolation between the two inventory level points where the values ranges between 0 and 1. By using the weight ratio the fixed order quantity policy changes to a base-stock policy as the inventory level increases.

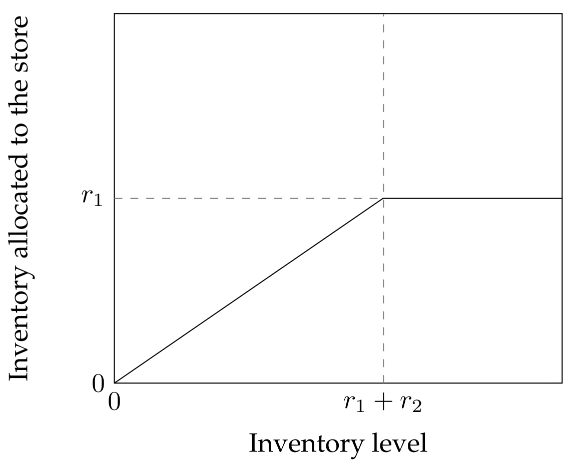
Figure 2.4(b) gives the algorithm for ordering, where first the parameters are derived followed by checking in which region the inventory level is. As no closed form exist for the inverse of the Normal distribution function, this takes a numerical procedure. Except for this calculation, the heuristic is able to find the order quantity fast.

2.5.2 Rationing Heuristic

The heuristic for rationing is based on operating the channels individually and trying to find the optimal quantity to allocate to each channel by considering the marginal costs and revenue for each additional product allocated to the two channels. Excess products the retailer has available are stored in the channel with the lowest inventory cost. The optimal quantity for each individual channel is found using the newsvendor model, however, the ordering cost is replaced by the holding cost. If the inventory level is below the optimal quantity of the individual channels, the decision on where to allocate a product is based on which channel gives the highest expected contribution for the product. G_i^{-1} denotes the inverse of the Poisson distribution function of either channel.

Figure 2.5(a) illustrates the structure found for the optimal rationing policy. In the figure the threshold inventory level where excess products are stored in the channel with the lowest inventory cost is indicated as $r_1 + r_2$. Thus r_1 is the optimal number of products available for the in-store customers taking into account the trade-off between cost and profit. Similarly, r_2 is the optimal number of products available for the online

(a)



(b)

Algorithm 3: Heuristic for rationing.

$r_1 = G_1^{-1}\left((c_p - (c_{h1} - c_{h2}))/p\right)$
$r_2 = G_2^{-1}\left((p - c_u) - c_{h2}/(p - c_u)\right)$
if $I \geq r_1 + r_2$ **then**
  $a = r_1$
else
  $i = 0, a = 0$
  while $i < I$ **do**
    $i = i + 1$
    $s = (i, 0)$
    if $\mathbb{E}R_{II}(s, a+1) > \mathbb{E}R_{II}(s, a)$ **then**
      $a = a + 1$
  end
end

Figure 2.5: Visualisation of the parameters of the rationing heuristic (a) with the algorithm (b).

customers. This inventory level point is found by calculating the optimal quantity for each individual channels. At inventory levels below this threshold, a trade-off is made between storing the product in the offline or online channel. This trade-off is based on which channel gives the highest expected contribution for an extra product.

Figure 2.5(b) gives the algorithm for rationing, where first the optimal quantity of the channels is derived as this takes a numerical procedure. The calculation of the highest expected contribution for the product is cumbersome however, unavoidable for the heuristic. Nevertheless, the heuristic for the rationing is able to find the rationing decision fast.

2.6 Performance of the Heuristics

The result of the heuristics is evaluated for the different problem instances described in Section 2.4.1. We evaluate the heuristics on three different performance measures, the goodness of fit of the policy, the optimality gap and the service level. The first performance is to evaluate how much the heuristics results resembles the optimal policy, the second performance indicator gives us an indication how good the heuristics performs on the different revenue and cost parameters. The last performance measures gives an indication on the performance of the fulfilment of the individual channels.

2.6.1 Goodness of fit

To evaluate the goodness of fit of the heuristic, a weighted root-mean-square-error (wRMSE) is used. RMSE sums the squared differences between the optimal actions and the actions set by the heuristic over all feasible states. We chose to use RMSE because it penalizes large deviations from the optimal decisions more than small deviations. Furthermore, we chose to use a weighted version of this measure because not all states are equaly likely to occur, and it is important that the heuristics fits well to the optimal policy in the states that are most predominant. To achieve this, we use weights set by the probabilities that the states occurs. These probabilities can be obtained from solving the steady state distribution by Markov chain analysis, or by only considering actions that are made during a simulation period. We apply simulation, and compute the wRMSE with weights set to the relative frequency that states are visited during a simulation. Similar as described in Section 2.4.3, we simulate the heuristics for J periods and calculate the wRMSE as follows:

$$\text{wRMSE} = \sqrt{\frac{\sum_{j=1}^{J} (\pi_j - k_j)^2}{J}} \tag{2.10}$$

where π_j is the optimal action to be taken in simulation period j and k_j is the action chosen by the heuristic in the same period. For the ordering action the simulation period $J = 100.000$ weeks and for the rationing action $J = 700.000$ days. By using a relatively long simulation period, we implicitly include the frequency of states occurring, as the error encountered for common states will be included many times in this weighted average. In Table 2.5 the wRMSE of both the ordering and rationing decision is presented.

From Table 2.5 it is observed that both the ordering and the rationing decision show low weighted RMSEs, where the maximum values are 1.699 and 1.031 respectively. For different lead times, it is observed that the weighted RMSE of ordering varies the most, where it increases with longer lead times. This can be explained by the fact that the heuristics tries to predict the quantity of products sold during lead time, and for shorter lead times this prediction is more accurate. For different review periods the weighted RMSE stays around 1.25.

The goodness of fit of the rationing decision is relatively stable for different review periods and lead times. The demand and cost parameters have the largest influence on

Table 2.5: Weighted RMSE of the heuristics during review period for all instances.

	Weighted RMSE	
Instance	Ordering	Rationing
$L = 1$	0.390	0.646
$L = 2$*	0.663	0.639
$L = 3$	1.066	0.681
$L = 4$	0.954	0.664
$L = 5$	1.272	0.645
$L = 6$	1.699	0.620
$L = 7$	1.548	0.658
$R = 2$	1.531	0.700
$R = 3$	1.678	0.761
$R = 4$	0.980	0.645
$R = 5$	0.998	0.677
$R = 6$	1.058	0.702
$R = 3, L = 3$	0.893	0.589
$R = 4, L = 4$	0.752	0.611
$R = 5, L = 5$	1.198	0.614
$R = 6, L = 6$	1.658	0.597
$\mu_1 = 2, \mu_2 = 2$	0.259	0.131
$\mu_1 = 4, \mu_2 = 4$	0.666	0.857
$\mu_1 = 2, \mu_2 = 6$	0.592	0.122
$c_u = 0$	0.592	0.649
$c_u = 20$	0.899	0.593
$c_{h1} = 2$	1.237	1.031
$c_{h1} = 3$	1.361	0.446
$c = 20$	0.980	0.501
$c = 40$	0.499	0.970

*Base test case with $R = 7, L = 2, \mu_1 = 6$, $\mu_2 = 3, p = 100, c_p = 30, c_u = 5, c_{h1} = 1$, $c_{h2} = 0.5$

the weighted RMSE, which is expected as these parameters are used in the rationing heuristic thus have the largest influence.

2.6.2 Optimality gap

Table 2.6 shows the profit of the MDP and the heuristics. Additionally, the gap of the revenue and all individual cost factors are included. From the results of the profit, it appears that the heuristics perform quite well and is very close to the optimum in all cases. The largest gap in profit is 0.064% for the instance in which $c_{h1} = 2$. The heuristics performance are the best for the instance in which $\mu_1 = 2, \mu_2 = 2$ with an gap of only 0.002%. These results correspond with the weighted RMSE as in these

Table 2.6: Optimality gap of the heuristics during review period for all instances.

	Profit			Revenue	Costs		
Instance	MDP	Heuristic	Gap (%)	Gap (%)	Inventory gap(%)	Ordering gap (%)	Shipment gap(%)
$L=1$	3626.63	3626.30	-0.009	0.031	0.858	0.020	-0.014
$L=2$*	3623.84	3623.42	-0.011	0.042	1.007	0.041	0.004
$L=3$	3621.15	3620.55	-0.017	-0.026	-0.171	-0.027	-0.089
$L=4$	3618.69	3618.07	-0.017	0.057	1.339	0.057	0.028
$L=5$	3616.36	3615.56	-0.022	0.079	1.934	0.061	0.063
$L=6$	3614.18	3613.05	-0.031	0.147	3.130	0.143	0.160
$L=7$	3612.09	3611.02	-0.030	0.106	2.355	0.106	0.107
$R=2$	1057.47	1056.99	-0.045	-0.175	-3.828	-0.176	-0.348
$R=3$	1579.53	1578.81	-0.046	-0.182	-3.606	-0.183	-0.326
$R=4$	2097.13	2096.78	-0.017	-0.061	-1.060	0.058	-0.146
$R=5$	2610.35	2609.49	-0.033	-0.060	-0.752	0.038	-0.146
$R=6$	3119.23	3118.46	-0.025	-0.055	-0.678	-0.048	-0.139
$R=3, L=3$	1577.80	1577.64	-0.010	-0.030	-0.332	-0.036	-0.284
$R=4, L=4$	2093.23	2092.93	-0.014	0.045	-3.253	0.508	-0.001
$R=5, L=5$	2604.00	2603.49	-0.020	0.076	1.992	0.071	0.052
$R=6, L=6$	3110.24	3109.30	-0.030	0.146	3.695	0.144	0.159
$\mu_1=2, \mu_2=2$	1762.99	1762.95	-0.002	0.008	0.166	0.008	0.007
$\mu_1=4, \mu_2=4$	3561.35	3560.84	-0.014	0.070	1.652	0.069	0.008
$\mu_1=2, \mu_2=6$	3507.54	3507.41	-0.004	0.016	0.399	0.015	0.017
$c_u=0$	3693.39	3692.97	-0.011	0.039	0.942	0.038	0.000
$c_u=20$	3415.46	3414.91	-0.016	0.068	1.340	0.066	0.165
$c_{h1}=2$	3542.67	3540.40	-0.064	0.312	5.043	0.311	0.041
$c_{h1}=3$	3467.23	3465.88	-0.039	0.188	2.447	0.186	0.052
$c_p=20$	4180.92	4180.47	-0.011	-0.055	-3.433	0.406	-0.107
$c_p=40$	3067.30	3066.60	-0.023	0.080	1.663	0.080	-0.015

*Base test case with $R=7, L=2, \mu_1=6, \mu_2=3, p=100, c_p=30, c_u=5, c_{h1}=1, c_{h2}=0.5$

instances they were also the smallest and largest.

From the performance results, it is observed that the heuristics are performing better for shorter lead times than for longer lead times, even though it is still close to optimal. As described above, the heuristics try to predict the quantity of products sold during lead time, which is more accurate for short lead times. This can also be seen by the cost of ordering, which increases with the lead time.

For short review periods, the performance results are the lowest. For short review periods, the amount of days with low inventory is relatively high compared with long review periods. The rationing decision influences the revenue and inventory cost, which can be seen as highest in the cases with short review periods. The high shipment cost gap indicates that more product are sold in the online channel than offline channel, while a product in the offline channel has a larger profit margin. The gap for profit at higher review periods remains around 0.020%.

For instances in which the review period is equal to the lead time, the gap is the lowest when they are both equal to three days. When decreasing or increasing the period the gap increases. The trend in gap is similar as discussed above, where longer lead time increase the gap and longer review period decreases the gap. However, the gap resulting from the sub-optimal rationing policy is more significant than the order policy, as the gap for short review period and lead time is larger than long review period and lead time. This can also be seen in the inventory cost, which gaps are significantly larger than the ordering gaps.

For different review periods and lead times, the heuristics are performing effective. Additionally the heuristics are tested for different demand distributions. Among all instances, the gap of the heuristics are the lowest for the different demand distributions.

The heuristics are also showing near-optimal results for instances with different cost parameters. Higher costs for handling online demand or higher costs for in-store inventory do increases the gap slightly. Having higher costs means that having sub-optimal rationing decisions decreases the profit more. Increasing the ordering cost decreases the gap of the heuristics. This can be explained by the approximation of the safety stock of the ordering heuristic. It assumes that a lost sale cost only consist of the profit of selling a product in the offline channel, not taking into account the possibility of the lost sale occurring in the online channel.

2.6.3 Service level

Next to profit an important performance indicator is the alpha service level, that is the fraction of time one ends a day with products still in stock (i.e., the non-stock out probability). We report these for every day and for each channel. In particular one is interested in the cycle service level, that is the product availability just before a replenishment arrives.

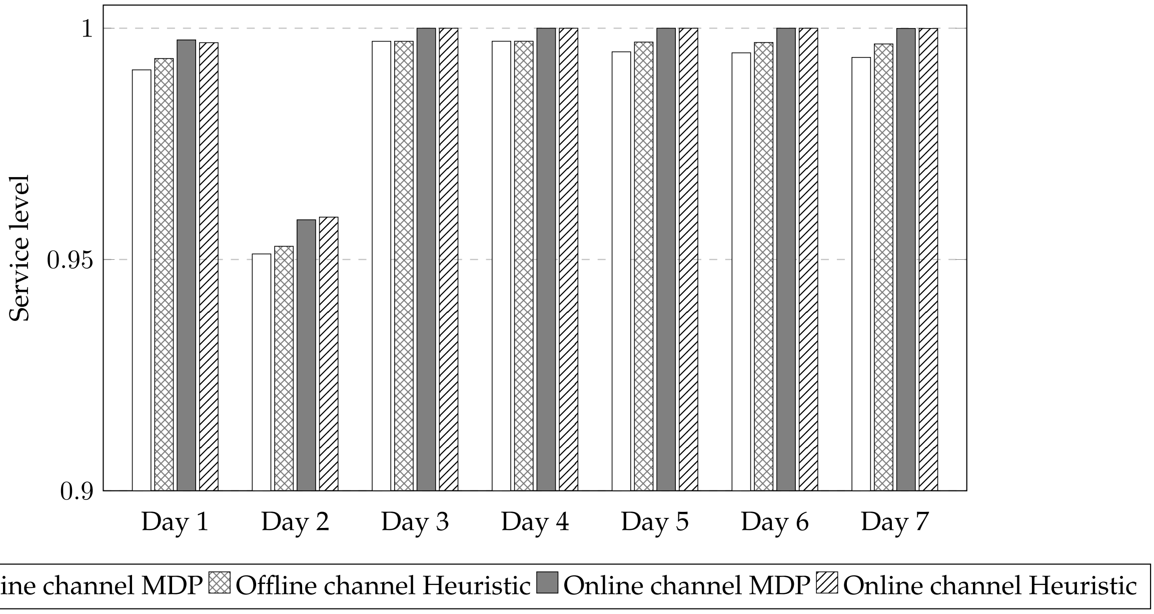


Figure 2.6: Average service level per day of each channel for the MDP and heuristic for the base case.

Service level per weekday

Figure 2.6 shows the average service level per day resulting from using the policy of the MDP and heuristics for each channel. The service level is calculated by the fraction of 100.000 simulated weeks, in which all demand can be met from stock assigned to that channel. The figure shows that demand in the online channel is fully served for the days following replenishment (day 3 till 7), and that there is some lost demand during those days in the offline channel. This is due to the rationing decision, and the holding costs being higher at the offline channel: it deems optimal not to meet the maximum possible demand of that channel. During the procurement lead time (day 1 and 2), the service level drops, as the retailer is experiencing lower inventory levels, and increased probabilities that they cannot satisfy all demand. Overall, the heuristic has a slightly higher service level than the optimal result. This is due to a slightly larger replenishment order, which happens mostly when the stock level is low upon ordering.

At high inventory levels, the two channels are not competing for products and the rationing decision can be reduced to a cost minimisation decision to determine the optimal base stock level for the offline channel. As having more products in the offline channel than the base stock level causes the expected marginal revenue to become less than the expected marginal cost, the optimal solution is to store excess products in the online channel. In extreme cases the offline channel might not have enough products to meet all demand. As the online channel holds all excessive stock it is capable to satisfy all demand in extreme cases. Both the policy of the MDP and the heuristic have this base stock level for the offline demand, therefore the service level is identical for day 3 and 4.

The day before replenishment (day 2), the trade-off between serving the two channels is more important due to lower inventory levels. Therefore, it is important that the heuristic is able to capture the trade-off between expected future sales and holding cost as good as possible. As seen in Table 2.5, the rationing decision has a low RMSE, indicating that the heuristic is capable of capturing the core trade-off.

The overall higher service level was also seen in Table 2.7, where the revenue is higher for the heuristic. The heuristic has an overall higher service level due to an average higher inventory level. The higher inventory level results from the larger replenishment orders, which was seen in Table 2.6 where the ordering gap is positive, indicating higher ordering costs for the heuristic. The average higher inventory level comes at the expense of higher inventory costs, which offset the higher revenue resulting in a lower profit for the heuristic.

Cycle service level

The service level on day l sets the cycle service level, which is an important performance indicator. Where Figure 2.6, focuses on the base case, in Table 2.7 we present the cycle service level for all instances. The cycle service level is the service level on the day just before replenishment. For most instances, the cycle service level is above 95%, only for the instances in which the holding cost is higher or the cost of the product is higher, the service level turns out to be lower. This can be explained by the fact that the cost of trying to satisfy demand is higher in these instances, and that it therefore is more profitable to not always satisfy demand. In settings where holding costs are much smaller, the fill rates may get close to one, and we observed that the performance of the heuristic is equally close or even closer to optimal.

Table 2.7 shows that the cycle service level is almost always slightly higher for the online channel than the offline channel, as the holding costs in the offline channel are larger. Due to the higher holding costs, fewer products are stored in the offline channel

Table 2.7: The cycle service level for all instances.

	Offline channel		Online channel	
Instances	MDP	Heuristic	MDP	Heuristic
$L = 1$	0.949	0.949	0.957	0.957
$L = 2$*	0.951	0.953	0.959	0.959
$L = 3$	0.951	0.946	0.960	0.954
$L = 4$	0.947	0.951	0.956	0.958
$L = 5$	0.949	0.955	0.956	0.961
$L = 6$	0.945	0.956	0.955	0.965
$L = 7$	0.946	0.954	0.954	0.961
$R = 2$	0.985	0.977	0.990	0.983
$R = 3$	0.979	0.968	0.985	0.974
$R = 4$	0.972	0.966	0.979	0.973
$R = 5$	0.965	0.959	0.972	0.965
$R = 6$	0.957	0.950	0.966	0.958
$R = 3, L = 3$	0.978	0.976	0.985	0.981
$R = 4, L = 4$	0.969	0.971	0.976	0.976
$R = 5, L = 5$	0.962	0.966	0.969	0.973
$R = 6, L = 6$	0.953	0.964	0.961	0.969
$\mu_1 = 2, \mu_2 = 2$	0.961	0.961	0.956	0.956
$\mu_1 = 4, \mu_2 = 4$	0.955	0.958	0.952	0.952
$\mu_1 = 2, \mu_2 = 6$	0.955	0.957	0.953	0.955
$c_u = 0$	0.950	0.952	0.961	0.960
$c_u = 20$	0.953	0.955	0.949	0.955
$c_{h1} = 2$	0.933	0.950	0.956	0.961
$c_{h1} = 3$	0.932	0.944	0.960	0.966
$c_p = 20$	0.959	0.953	0.966	0.959
$c_p = 40$	0.938	0.941	0.949	0.948

*Base test case with $R = 7, L = 2, \mu_1 = 6, \mu_2 = 3, p = 100$, $c_p = 30, c_u = 5, c_{h1} = 1, c_{h2} = 0.5$

resulting in a lower service level.

The service level remains relatively stable for different lead times, as for the offline channel it is around 95% and for the online channel between 95-96%. Though not visible here, we have observed that for a longer lead time the optimal policy resembles less a base stock policy. This is explained by the fact the actual inventory level gives less information on the inventory level when products are delivered after a longer lead time. The proposed heuristic closely approximates the optimal ordering policy by a combination of a constant order policy and a base stock policy, as discussed in Section 2.6.1. Thus we achieve comparable service levels.

When increasing the review period, the cycle service level decreases. Longer review periods increase the uncertainty of demand thus the chance of hitting a stock out before replenishment increases as the retailer is less responsive to low stock levels. This is also seen in the instances in which the review period is equal to the lead time.

For most instances, the MDP results in a lower service level than the heuristic. The heuristic often has higher replenishment orders than the MDP, resulting in the higher cycle service levels at the expense of higher inventory and ordering costs (as was shown in Table 2.6. However, for short review periods, the heuristic orders fewer products than the MDP resulting in lower cycle service levels.

2.7 Conclusion and Discussion

Brick-and-mortar stores are increasingly adopting online shopping channels in their portfolio, where in-store inventory is often used to fulfil both in-store demand as well as online orders. As the retailer serves demand of multiple channels by one inventory, this study proposes to ration the inventory of the retailer among the channels. The rationing is suggested as solution to the negative effects that occur when using in-store inventory for the fulfilment of online demand.

The rationing is characterised by the trade-off of serving in-store or online customers, where the profit of selling a product in-store is higher than through the online channel but the inventory cost of the online channel is lower. The proposed model encompasses this core trade-off experienced by an omni-channel retailer. The underlying cost structure and demand of both channels influence this decision. The products are present at the same location, thus the retailer can make the rationing decision on a daily basis.

This paper contributes to the academic literature on omni-channel retailing by providing a novel exact method as well as heuristics for the integration of the replenishment and rationing decision. Previous research on dynamic rationing assumes zero lead times, no lost sales, and static order policies which does not capture the characteristics of an omni-channel retailer. To capture these characteristics, we study an omni-channel retailer with a dynamic order policy, a deterministic lead time, and lost sales. The problem is formulated as an MDP that was solved through value iteration to get an exact solution. The structure of the optimal rationing and ordering policies are numerically investigated and captured in heuristics.

Based on our model and numerical results, we conclude that the optimal ordering policy consists of a fixed order quantity and an order-up-to level. At low inventory levels, a fixed order quantity is optimal, whereas at high inventory levels an order-up-to level is optimal, this is due to the uncertainty of demand during the procurement lead time. For the rationing policy, a maximum threshold level for high inventory levels is found, which is also proven analytically. At low inventory levels there is a trade-off on whether to store products in-store or offline. Based on these insights, heuristics are developed and tested through extensive numerical tests. The heuristics are compared with the optimal policy on its goodness of fit, optimality gap and service level. The heuristics show near-optimal results, concluding that the heuristics can be effective in encompassing the complexity of dynamic rationing. Furthermore, the heuristics are able to closely approximate the optimal ordering policy by mixing a base stock policy and a constant order policy.

Even though the model studied in this paper captures the core trade-offs presented in omni-channel retailing, several potentially complicating factors were not taken into account, and could be directions for further research. More specifically, it could be in the exploration of a network of multiple omni-channel retailer locations who can fulfil online demand, extending the rationing decision with the allocation of online orders to different locations. Additionally, retailers might want to steer customers more towards the store. The retailer might make the decision to offer certain products only in-store and perhaps release them later online. Loyal store customers often have higher customer lifetime value and should therefore be rewarded. In this research, we considered the setting where the rationing decision is made at the beginning of the day. However, the retailer might want to rethink the rationing decision throughout the day. With the advancement of RFID technology the retailer has better insights into their inventory position and thus is able to make this decision accurately and more frequent. It is interesting to investigate the influence of the timing and frequency of the rationing decision on the profit and service levels of the omni-channel retailer. The current model can easily be extended to this setting by changing the number of sub-periods in the model.

This research contributes to such a direction by offering heuristics that are fast and have the potential to be upgraded to these more complex problem settings. However, it should be noted that use of our heuristics in practice would require the retailer to regularly update demand parameters throughout the selling period as these might change over time. In our paper we assumed lost sales as replenishment happens only weekly and customers may be impatient. By adopting the lost sales assumption, we prevent stock outs as much as possible as they would result in lost profit. A mixture of lost sales and backordering could be incorporated in the framework we present, e.g. if a percentage of online customers might be willing to wait for the fulfilment of online orders. Adding a backlog decision will further increase the computational complexity. Based on previous research (e.g. Benjaafar et al., 2010; Fadıloğlu & Bulut, 2010), we expect the amount of backorders to have a limit and that the decision of when to backorder will dependent on the state of the system. Additional research could be interesting to investigate the influence of product returns in the current context. Online sales have become challenging for retailers due to the large quantity of sold goods being returned. This return flow influences the inventory of the retailer, thus influencing the rationing and ordering policies. The addition of returns would however significantly increase the dimensionality of the MDP, as the state of the model would have to include the numbers of sold products (in the last days or weeks) that will be potentially returned in the coming days or weeks. The curse of dimensionality would result in unsolvable MDPs for most relevant problem sizes. The development of effective heuristic solutions would then also be highly relevant for this problem.

Appendix 2.A Derivation of Inventory Base Stock Level

We can prove that there is an optimal quantity to be stored in-store in case of excessive stock. We have the following equation to calculate the expected contribution in state s_t when taking action a:

$$\begin{aligned}\mathbb{E}C_{\text{II}}\left(s_t,a\right) = p &\left(\sum_{d<a} d\cdot P_1\left(d\right) + \sum_{d\geq a} a\cdot P_1\left(d\right)\right)\\ &+\left(p-c_u\right)\left(\sum_{d<I_t-a} d\cdot P_2\left(d\right) + \sum_{d\geq I_t-a}\left(I_t-a\right)\cdot P_2\left(d\right)\right)\\ &-\left(c_{h1}\cdot a + c_{h2}\cdot\left(I_t-a\right)\right)\end{aligned}$$

We add the following term:

$$p\sum_{d\geq a+1} d\cdot P_1\left(d\right) - p\sum_{d\geq a+1} d\cdot P_1\left(d\right) + \left(p-c_u\right)\sum_{d\geq I_t-a+1} d\cdot P_2\left(d\right) - \left(p-c_u\right)\sum_{d\geq I_t-a+1} d\cdot P_2\left(d\right)$$

We know that the following two conditions will always hold

$$\sum_{d<a} d\cdot P_1\left(d\right) + \sum_{d\geq a+1} d\cdot P_1\left(d\right) = \mu_1 \quad \text{and} \quad \sum_{d<a} d\cdot P_2\left(d\right) + \sum_{d\geq a+1} d\cdot P_2\left(d\right) = \mu_2$$

We can therefore get to the following equation after adding and subtracting:

$$\begin{aligned}\mathbb{E}C_{\text{II}}\left(s_t,a\right) \;=\; & p\cdot\mu_1 + p\sum_{d\geq a+1}\left(a-d\right)\cdot P_1\left(d\right) + \left(p-c_u\right)\cdot\mu_2\\ &+\left(p-c_u\right)\sum_{d\geq I_t-a+1}\left(I_t-a-d\right)\cdot P_2\left(d\right) - \left(c_{h1}\cdot a + c_{h2}\cdot\left(I_t-a\right)\right)\end{aligned}$$

To find the optimal action a that maximises the expected contribution $\mathbb{E}C\left(s_t,a\right)$, we take the approximate derivative:

$$\Delta\mathbb{E}C_{\text{II}}\left(s_t,a\right) = \mathbb{E}C_{\text{II}}\left(s_t,a+1\right) - \mathbb{E}C_{\text{II}}\left(s_t,a\right)$$

After subtracting and rewriting we get the following equation:

$$\Delta\mathbb{E}C_{\text{II}}\left(s_t, a\right) = p\sum_{d\geq a+1} -P_1\left(d\right) + \left(p - c_u\right)\cdot\sum_{d\geq I_t - a+1} P_2\left(d\right) + \left(c_{h1} - c_{h2}\right)$$

We know that $\sum_{d\geq a+1} P_1\left(d\right) = 1 - \sum_{d\leq a} P_1\left(d\right)$ and $\sum_{d\geq I_t-a+1} P_2\left(d\right) = 1 - \sum_{d\leq I_t - a} P_2\left(d\right)$

We can therefore rewrite the approximate derivative as follows:

$$\Delta\mathbb{E}C_{\text{II}}\left(s_t, a\right) = -p + p\cdot\sum_{d\leq a} P_1\left(d\right) + \left(p - c_u\right) - \left(p - c_u\right)\cdot\sum_{d\leq I_t - a} P_2\left(d\right) + \left(c_{h1} - c_{h2}\right)$$

If all possible demand in the online channel is satisfied due to excess inventory being available to this channel, we have $\sum_{d\leq I_t-a} P_2\left(d\right) = 1$. This assumes $h_1 > h_2$, as otherwise the excess inventory would be allocated to the offline channel. In this situation, the approximate derivative is reduced to the following:

$$\Delta\mathbb{E}C_{\text{II}}\left(s_t, a\right) = -p + p\cdot\sum_{d\leq a} P_1\left(d\right) + \left(c_{h1} - c_{h2}\right)$$

In order to find the maximum value of $\mathbb{E}C_{\text{II}}\left(s_t, a\right)$, we find the largest value for a for which $\Delta\mathbb{E}C_{\text{II}}\left(s_t, a\right)$ is negative:

$$\begin{aligned}\Delta\mathbb{E}C_{\text{II}}\left(s_t, a\right) \leq 0 \quad &\Longrightarrow \quad -p + p\cdot\sum_{d\leq a} P_1\left(d\right) + \left(c_{h1} - c_{h2}\right) \leq 0 \\ &\Longrightarrow \quad \sum_{d\leq a} P_1\left(d\right) \leq \frac{p - \left(c_{h1} - c_{h2}\right)}{p}\end{aligned}$$

Thus we get the following equation that gives us the optimum amount of products to hold in-store:

$$P_1(d \leq a) \leq \frac{p - \left(c_{h1} - c_{h2}\right)}{p}$$

The inventory base stock level a can then be easily derived from the inverse probability function of P_1.

Chapter 3

Modelling the Influence of Returns for an Omni-channel Retailer

This chapter is published as:

Goedhart, J., Haijema, R. & Akkerman, R., (2023). Modelling the influence of returns for an omni-channel retailer. *European Journal of Operation Research*, *306*, 1248-1263

Abstract

More brick-and-mortar retailers open an online channel to increase sales. Often, they use the store to fulfil online orders and to receive returned products. The uncertain product returns however complicate the replenishment decision of a retailer. The inventory also has to be rationed over the offline and online sales channels. We therefore integrate the rationing and ordering decisions of an omni-channel retailer in a Markov Decision Process (MDP) that maximises the retailer's profit. Contrary to previous studies, we explicitly model multi-period sales-dependent returns, which is more realistic and leads to higher profit and service levels. With Value Iteration (VI) an exact solution can only be computed for relatively small-scale instances. For solving large-scale instances, we constructed a Deep Reinforcement Learning (DRL) algorithm. The different methods are compared in an extensive numerical study of small-scale instances to gain insights. The results show that the running time of VI increases exponentially in the problem size, while the running time of DRL is high but scales well. DRL has a low optimality gap but the performance drops when there is a higher level of uncertainty or if the profit trade-off between different actions is minimal. Our approach of modelling multi-period sales-dependent product returns outperforms other methods. Furthermore, based on large-scale instances, we find that increasing online returns lowers the profit and the service level in the offline channel. However, longer return windows do not influence the retailer's profit.

3.1 Introduction

The retail sector is changing drastically with the rise of online shopping. As customers are shopping more online, traditional brick-and-mortar stores are also changing their business strategy by opening online shopping channels. This integration of offline and online shopping channels is referred to as omni-channel retailing (Verhoef et al., 2015). Omni-channel retailing provides the customer a uniform shopping experience, in which goods can be inspected, bought, and returned through all available shopping channels. However, customers that order products online are only able to physically inspect their goods after delivery. Therefore, products that are displayed online might not satisfy customer expectations or are bought impulsively, online ordered products are often returned (Abdulla et al., 2019).

The return flow of online ordered products has become a significant issue for many retailers. For instance, an online fashion retailer reported return percentages between 13% and 45% (de Leeuw et al., 2016). As the returned products are often resalable they should be accounted for in inventory management (Radhi & Zhang, 2019). Due to the uncertainty in the quantity and timing of returned products, retailers might end up with an excessive stock if they do not consider the return flow (Bernon et al., 2016; Xu & Jackson, 2019). Even though retailers sometimes try to reduce returns (e.g., through stricter return windows or return fees), in many cases retailers are providing increasingly lenient return policies to increase customer satisfaction. Therefore, retailers need to adapt their inventory management to the growing return flow as the handling of returned products is important to reduce inventory related cost.

Customers often prefer the ability to return an online purchase in-store. Brick-and-mortar store returns are free of charge for customers and do not require repackaging, whereas via mail the customer has to deal with packaging and often has to pay a shipping fee. Additionally, the customer can get a direct replacement in the store or immediate refund (Wollenburg et al., 2018). It can also be more profitable for retailers to encourage customers to return products to the store instead of shipping via mail (Nageswaran et al., 2020). With using the brick-and-mortar store for returns they can steer the customer towards exchanging their product or towards buying another product (Tarn et al., 2003). Furthermore, the retailer can use the brick-and-mortar store to inspect returns, thus not accepting invalid or unwanted returns (de Leeuw et al., 2016). Due to such gatekeeping, returns are processed faster, which is important to sustain the value of the product as they can quickly be added back to the store inventory (Hübner et al., 2015). Allowing returns from online sales to be returned in-store is referred to as cross-channel returns (Radhi & Zhang, 2019). However, the retailer needs to account for these returns in their ordering decision as not accounting for the returns could result in imbalanced inventory levels and significant revenue loss (Chen & Bell, 2012; Hu et al., 2019).

The role of the store for an omni-channel retailer has becoming increasingly important, where the store can operate as a fulfilment centre for online orders, a pick-up point, a place to handle returns, or an information channel (Mou et al., 2018; Hübner et al., 2022). Shipping online ordered products from stores is referred to as a ship-from-store strategy (Agatz et al., 2008). The strategy has several advantages and disadvantages for the retailer. Advantages are for instance lower inventory levels, higher turnover rates, and shorter delivery distances (Bayram & Cesaret, 2021; Jalilipour Alishah et al., 2015), and examples of disadvantages are negative in-store customer experiences due to store personnel picking orders and inaccurate inventory positions. To mitigate these negative side-effects managerial studies suggest to reserve part of the brick-and-mortar store inventory for the online demand (Hobkirk, 2015; ENC, 2016). This reservation of part of the inventory is referred to as rationing the inventory across the shopping channels.

By rationing an inventory across channels, a trade-off is made. Storing products in the offline channel is more costly due to expensive shelve space in the store. However, online channels typically have a reduced profit per product due to the cost of online fulfilment and the probability of a product being returned. Little research has been conducted on the trade-off a retailer has to make between the offline and online channel in the context of using the store assets for the online channel. In this paper, we therefore study how a retailer can utilise their brick-and-mortar store to handle the online fulfilment as well as the potential returns. The objective of this study is to identify an optimal replenishment and rationing policy for an omni-channel retailer taking into account the return flow of the online ordered products.

The integration of returns in the ordering and rationing decision complicates the studied problem, as predictions for returns have to be considered in modelling the decision problem. Clearly, returns will depend on historical sales. However, keeping track of historical sales can easily make modelling approaches intractable. Therefore, historical sales data is often approximated by aggregating detailed historical sales data, so the information can still be used in some way to make better decisions. If retailers would not take into account potential returns at all, they would end up with excessive stock. Additionally, retailers can choose to manage their inventory such that the potential returns from the online sales channel are considered in setting ordering and rationing policies. For instance, the rationing decision can be used to reserve products for future in-store customers who are more profitable when inventory positions and outstanding orders are low.

This research contributes to the literature on omni-channel retailing by showing how returns affect the retailer's profit and inventory management. More specifically, we first provide a model that explicitly considers multi-period sales-dependent returns in the inventory management of an omni-channel retailer, based on a Markov Decision Process (MDP) formulation. Second, as the MDP might become too large to solve large-scale instances with value iteration (VI) to obtain an exact solution, we demonstrate how Deep Reinforcement Learning (DRL) can be used to solve the problem and obtain an approximated solution. Third, we compare the multi-period sales-dependent return MDP with other methods of modelling returns to gain insight in the importance of incorporating historical sales in decision-making. Fourth, based on our numerical results, we provide managerial insights on how an omni-channel retailer should cope with returns, as well as general insights on the use of DRL in the context of retail operations inventory management.

The remainder of this paper is structured as follows. Section 3.2 presents related research on return strategies for omni-channel settings and inventory management. In Section 3.3, we further outline the decision problem and formulate it as a MDP. Section 3.4 presents the implementation of the DRL algorithm to the studied problem. In Section 3.5, the performance of the DRL policy compared to the optimal solution is investigated for a wide range of instances on different performance measures. Additionally, the importance of including historical sales data for decision-making is investigated by comparing it with other methods of modelling returns. In Section 3.6 we derive managerial insights from large-scale instances. Section 3.7 concludes the research and discusses future research opportunities.

3.2 Literature Review

Our work is related to the literature on omni-channel retailing as well as the literature on inventory management with return flows. Below, we briefly address related work on omni-channel retailing, with a focus on returns in an omni-channel context and the role of the store in omni-channel retailing. This is followed by a discussion of the literature on different methods for modelling return flows in inventory management.

3.2.1 Omni-channel retailing

The study of return management in an omni-channel context is an understudied problem (Bernon et al., 2016; Xu & Jackson, 2019; Muir et al., 2019). Hübner et al. (2022) also mention that the current body of literature does not close the gap between inventory management and returns in an omnichannel context. However, for retailers, return management is becoming increasingly important, especially due to the increase in the return flow originating from online orders. Retailers offer lenient return policies to relieve customer shopping risk and increase demand. However, different return policies have different effects. Generous return policies increase demand, whereas longer return windows and exchange leniency influence return percentages (Janakiraman et al., 2016). Ketzenberg et al. (2020) mention that lenient return policies have resulted in customers exploiting the retailers policies.

The current body of literature on return management mostly focuses on only online retailers. For an overview of this literature, we refer to the recent comprehensive review by Abdulla et al. (2019). Most modelling research around returns is about return policies, and often only focuses on single-period return windows. Abdulla et al. (2019) also conclude that opportunities for future research lies in analysing how operational decisions regarding returns can be made to improve retailers performance.

The strategic side of handling returns by brick-and-mortar stores has been extensively studied, where the focus is often on whether stores should be used for handling returns or not (e.g. Jin et al., 2020; Mandal et al., 2021; Gao et al., 2022). However, the operational side of handling the returned products in-store has only been studied to a limited extent (Hübner et al., 2016a; Ishfaq et al., 2016). Here, one of the main issues is the re-balancing of inventory (Bernon et al., 2016). Muir et al. (2019) and Radhi & Zhang (2019) investigate how same- and cross-channel returns influence order policies, and conclude that leveraging stores for returns improves service levels as returned products can be resold quicker. Dijkstra et al. (2019) investigate how to re-balance the cross-channel returns across the stores or online fulfilment centre of the retailer.

3.2.2 Inventory management with return flows

Modelling product return flows is complicated by the interaction between inventories, sales, and returns. Return flows depend on historic sales, while current sales is limited by the inventory, which is in turn influenced by the returns. In the literature, we find two different ways to model product returns: (i) product returns are independent of demand, (ii) and product returns are dependent on demand.

By assuming that returns are independent of demand, return flows can be modelled as exogenous flows. Fleischmann et al. (2002) and Feinberg & Lewis (2005) mention that in such cases, the problem comes down to a variant of an inventory model with positive or negative net demand. However, Kiesmüller & Van der Laan (2001) show that neglecting the dependency between demand and return results in poor performance of inventory policies. Yet, Zerhouni et al. (2013) mention that ignoring the dependency between demand and return increased costs only minimal in their study, which can be attributed to the long return window they considered, damping the effect of demand fluctuations. The study setting is based on de Brito & Dekker (2003), who also mention that long return windows is an assumption that does not hold for most retail settings. Cases in which the return time is long, such as certain remanufacturing systems can neglect the dependency without much impact on performance (Fleischmann & Minner, 2004), due to the damping effect mentioned above.

Fleischmann & Kuik (2003) discuss that modelling returns dependent on demand is difficult, as such dependence spans across multiple periods. Therefore, early work on returns is focused on modelling the dependency only within the same period or across one period, as this reduces the complexity (DeCroix, 2006). However, such an assumption is often too harsh as approximating returns within at most one period results in sub-optimal policies. Having the returns span across multiple periods increases the modelling complexity significantly as it grows with the quantity of products sold and the number of periods considered for the return window. Benedito & Corominas (2013) consider a remanufacturing system in which the products are always returned, either when their lifetime has exceeded or are broken down during their lifetime. They do include a dependency across multiple periods and show that finding the optimal policy for a small example is computationally feasible, but for longer return windows they propose an approximated MDP where returns are independent of sales. Ambilkar et al. (2022) recently also emphasised that an interesting research direction is using machine learning techniques to solve the intractable return management problem.

3.2.3 Contribution

We conclude from the literature that modelling multi-period sales-dependent return flows is difficult. The complexity of modelling approaches typically grows exponentially with the length of the return window and the number of products sold in that window and thus easily becomes intractable. However, including this dependency of returns with sales is important to ensure optimal policies. Largely due to this complexity, the influence of returns on managing inventories for an omni-channel retailer is an understudied subject. Whereas most previous literature focuses on the strategic decision of whether or not to handle returns in-store, we specifically aim to study the impact of handling returns on operational decision making. We contribute to the existing literature by studying the inventory management of an omni-channel retailer that uses their store inventory to fulfil both offline and online demand. The products ordered online have a probability of being returned within a given return window, spanning over multiple selling periods. To reduce model complexity, and still capture some of the temporal dynamics, we introduce a modelling approach to address the complexity of return dependence across multiple periods.

3.3 Markov Decision Process

We study the setting of an omni-channel retailer with a physical store and an online channel. The retailer has two decisions to be made: an ordering decision and a rationing decision. We consider a setting in which a period consists of T sub-periods. We assume the retailer places an order at the beginning of a period, which is replenished after L sub-periods. This could for instance reflect a weekly ($T = 7$) ordering decision with a delivery after two days ($L = 2$). This ordering decision q occurs at the beginning of the first sub-period. Additionally, at the beginning of each sub-period, a rationing decision a for the two channels is made, deciding on the inventory levels that are reserved for each channel. These inventories are used to fulfil the demand of the channels, and no substitution between channels takes place. When a customer faces a stock-out in their preferred channel, the demand is assumed to be lost which is a com-

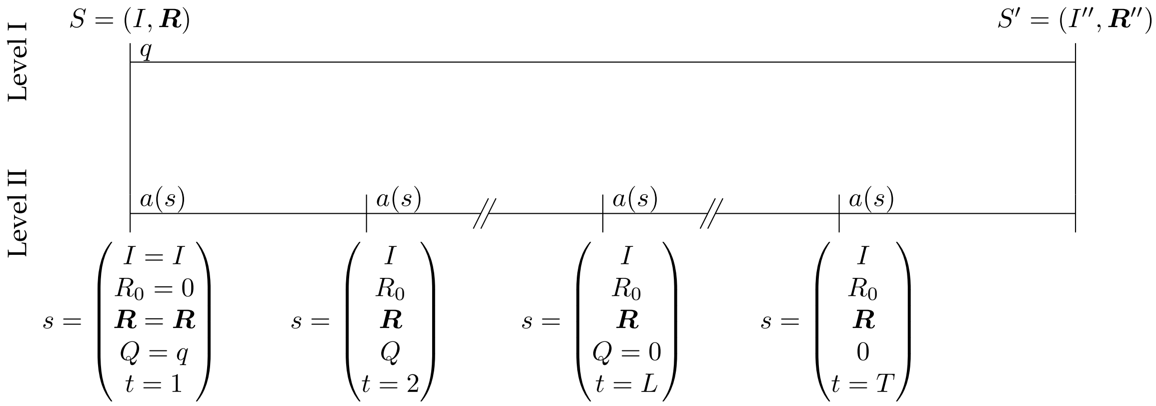


Figure 3.1: Visualisation of the states and actions in one period.

mon assumption in retail environments. The customer has a limited return window of M periods in which they can return the product. Throughout the sub-periods, products that were sold in previous periods in the online channel can be returned.

At the end of each sub-period, the remaining inventories of the two channels, together with the products returned during that sub-period, are used for fulfilment the next day. The retailer keeps track of the products sold in the online channel, as these can lead to returns. We assume that products sold through the offline channel are not returned, as product returns are mostly a characteristic of online sales.

3.3.1 Modelling approach

The problem described above can be formulated as an MDP. As the problem consists of decisions and state transitions on different time intervals, we formulate it as a Hierarchical MDP as described in Kristensen (1988). Describing the studied problem as a hierarchical MDP allows for convenient notation as action spaces and state transitions will not be dependent on the sub-period. The hierarchy of decisions in the MDP is illustrated in Figure 3.1. At level I of the MDP the replenishment decision is made, and at level II of the MDP the rationing decision is made in each sub-period. The selling process and possible returns are also included in level II for each sub-period.

The objective of the MDP is to maximise profit, consisting of the sales revenue and the different costs. These costs include the costs of fulfilment, handling returns, product, and holding costs for the individual channels. In an omni-channel setting, the price of the product is equal in both channels. We keep track of the number of products sold online each sub-period however, aggregate them on the time scale of level I to reduce complexity. As the retailer has a maximum time window for customer to return their products, we only keep track of products that are sold within M periods. Furthermore, we assume that products are not returned within the period the product is sold as customers do not instantly return a product after delivery.

At level I of the MDP, the state S consists of the inventory position I and the quantity of unreturned products sold in the online channel up to M periods ago $\boldsymbol{R} = (R_1, ..., R_M)$, together forming the state $S = (I, \boldsymbol{R})$. We use the bold notation to indicate it as a vector. At this level, the replenishment decision q is made. At level II of the MDP, the state s consists of the inventory position I, the number of products sold in the online channel so far this sub-period R_0, unreturned products from the previous periods $\boldsymbol{R}$, the outstanding replenishment order Q, and the sub-period t. At this level, the rationing decision a is made each sub-period, where a is the number of products stored in the offline channel. At the beginning of the first sub-period the information of the state at level I is used to construct the state at level II: $s = (I, R_0, \boldsymbol{R}, Q, t)$. The inventory state I and $\boldsymbol{R}$ are identical from the state at level I. For the first sub-period R_0 is set to zero, and Q is set to the replenishment decision of level I q. For all subsequent sub-periods, the number of online products sold in the previous sub-periods is added to R_0.

When the replenishment occurs at sub-period $t = L$, the outstanding order quantity is set to zero for the rest of the period, i.e. $s = (I, R_0, \boldsymbol{R}, Q = 0, t = L)$. At the end of all sub-periods, the total online sales of the period is given by R'_0, and the remaining number of unreturned products sold in the channel m periods ago by R'_m. This state information of level II is passed back to level I to form the state $S' = (I'', \boldsymbol{R}'')$. Where I'' is thus the closing inventory of the period, and $\boldsymbol{R}''$ is the sold products in the online channel that are not returned, as these are from the previous period they are given as $R''_1 = R'_0, R''_2 = R'_1, ..., R''_M = R'_{M-1}$.

The uncertainty in the MDP comes from both the demand and the product returns. The demand is modelled according to a discrete distribution, with $P_i\left(d_i\right)$ indicating the probability that demand of channel i is d_i with $i \in \{1 = \text{offline}, 2 = \text{online}\}$ and $d_i \in \{0, 1, \ldots, D_i\}$. D_i indicates the maximum possible demand of the channel. The returns are modelled according to M Binomial distributions given as $B\left(r_m\right)$ with parameters ρ and R_m, where ρ is the probability a single product is returned, and R_m is the pool of returnable items, which is the total sales of m periods ago that has not been returned yet. As we only assume products are returned that are sold between 1 and M

periods ago, $m \in \mathcal{M}$ with $\mathcal{M} = \{1, 2, \ldots, M\}$. We assume that the return probability is independent of the period in which the product is sold. If there would be a reason to model these probabilities differently, ρ could be made dependent on m as the binomial distributions of unreturned products sold in period m are independent.

3.3.2 Model

State space

The state space of the inventory state at level I is given by $I \in \mathcal{I}$ with $\mathcal{I} = \{0, 1, \ldots, (T + L) \cdot \sum_i D_i + M \cdot T \cdot D_2\}$. The inventory state is limited by the maximum expected demand and the returns of the last M periods. We assume the retailer will never order more than the maximum expected demand over the period plus lead time. However, products sold online over the last M periods ago could theoretically all be returned and added to the inventory. The maximum inventory level is therefore given by $(T + L) \cdot \sum_i D_i + M \cdot T \cdot D_2$. The state space of the unreturned product sold m periods ago is limited by the maximum amount of expected online sales in a period: $\boldsymbol{R} \in \mathcal{R}$ with $\mathcal{R} = \{0, 1, \ldots, T \cdot D_2\}$.

The state space of the inventory at level II is equal to the state space at level I, $I \in \mathcal{I}$. The state space of the order quantity is limited by the maximum expected demand of the review period minus the inventory position: $Q \in \mathcal{Q}(I)$ with $\mathcal{Q}(I) = \{0, 1, \ldots, T \cdot \sum_i D_i - I\}$. The state space of the sold product m periods ago is equal to that at level I, $\boldsymbol{R} \in \mathcal{R}$.

The state space of the products sold in the current period is dependent on the sub-period, as it is limited by the maximum amount of expected online sales per sub-period: $R_0 \in \mathcal{P}(t)$ with $\mathcal{P}(t) \in \{0, 1, \ldots, t \cdot D_2\}$.

Action and action space

At level I of the MDP, the order quantity q is determined at the beginning of the period, and is limited by the order quantity state: $q \in \mathcal{Q}\,(I) = \{0, 1, \ldots, T \cdot \sum_i D_i - I\}$.

At level II of the MDP, a rationing decision a is made every sub-period t. This rationing decision decides how many products are stored in the channels. The action space of the rationing decision is dependent on the current inventory position: $a \in \mathcal{A}(I)$ with $\mathcal{A}\,(I) = \{0, 1, \ldots, I\}$. The rationing decision is made at the beginning of the sub-period and is not revised throughout the sub-period.

State transitions

At level I of the MDP the state transitions every period, while at level II of the MDP the state transitions every sub-period. At level II of the MDP the state transition from $s = (I, R_0, \boldsymbol{R}, Q, t)$ to $s' = (I', R_0', \boldsymbol{R}', Q', t+1)$ occurs. The transition of the inventory I to I' is dependent on the demand, rationing decision, returns, and replenishment:

$$I' = (a - d_1)^+ + ((I - a) - d_2)^+ + \sum_{m=1}^{M} r_m + \delta\left(t = L\right) \cdot Q \tag{3.1}$$

Where $x^+ = \max(0, x)$ and $\delta(x)$ denotes the Kronecker delta, which gives the value 1 if $x =$ True, otherwise 0. The first term of the equation refers to the demand satisfied in the offline channel, in which the inventory level is a and the demand occurring d_1. The second term refers to the demand satisfied in the online channel, in which the inventory level is $I - a$ and the demand occurring d_2. The third term of the equation refers to the returned products, which is the sum of the products returned that are sold between 1 and M periods ago. The last term refers to the replenishment, which occurs at $t = L$, which is at the beginning of sub-period L.

The transition of the number of products sold online in the current period R_0 to R_0' is dependent on the number of product sold online the previous sub-period and the products sold online in the current sub-period:

$$R_0' = R_0 + \min\left(d_2, I - a\right) \tag{3.2}$$

Where the product sold online in the current sub-period are added to those where already sold in the subsequent sub-period. The transition of unreturned products m periods ago, R_m to the next state, R_m' is dependent on the quantity of products being returned r_m. Only unreturned products sold more than 1 period ago can be returned, with a maximum of M periods.

$$R_m' = R_m - r_m \qquad \forall m \in \mathcal{M} \tag{3.3}$$

The transition of the replenishment quantity Q to Q' depends on the sub-period. If the replenishment has yet to occur, Q remains constant, otherwise it is set to zero:

$$Q' = \begin{cases} Q & \text{if } t < L \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

At level I of the MDP the state only transitions at the end of the period, where the information of the state at level II is obtained to form the new state of level I: $S' = (I'', \boldsymbol{R}'')$. I'' is the closing inventory of the period, which is obtained after the last inventory state transition of level II of the MDP when $t = T$.

The state transition of the number of unreturned products for level I is obtained after the last state transition of R_0 and $\boldsymbol{R}$ from level II of the MDP when $t = T$. As we keep track of the periods ago a products was sold and the period has ended, unreturned products that were returned $m-1$ periods ago are now m periods old. Unreturned products older than M periods are not eligible to be returned anymore, thus we do not keep track of these items.

$$R''_m = R'_{m-1} \quad \forall m \in \mathcal{M} \tag{3.5}$$

Expected immediate reward

The goal of the MDP is to maximise its profit, which consists of the revenue, and several costs. The expected immediate reward at level I of the MDP only consists of the ordering cost and is given as follows:

$$\mathbb{E}C_{\mathrm{I}}\left(S, q\right) = -q \cdot c_p \tag{3.6}$$

In which c_p is the procurement cost of the product. The expected immediate reward per sub-period at level II of the MDP consists of revenue sold from selling products, online fulfilment cost, holding costs and handling costs of returns:

$$\mathbb{E}C_{\mathrm{II}}(s,a) = \begin{cases} p\left(\sum_{d<a} d \cdot P_1(d) + \sum_{d \geq a} a \cdot P_1(d)\right) & \text{(3.7a)} \\ + (p - c_u)\left(\sum_{d<I-a} d \cdot P_2(d) + \sum_{d \geq I-a} (I-a) \cdot P_2(d)\right) & \text{(3.7b)} \\ - (c_{h1} \cdot a + c_{h2} \cdot (I-a)) & \text{(3.7c)} \\ - (p + c_r) \sum_{m=1}^{M} R_m \cdot \rho & \text{(3.7d)} \end{cases}$$

The first term (3.7a) of the expected immediate reward consists of the revenue of the offline channel given price p. The same sales price is applied for the online channel, however, a shipment cost of c_u is incurred as seen in the second term (3.7b). As customers are expecting free shipping, the retailer has to incur the cost. The third term (3.7c) is the holding cost of the offline and online channel, c_{h1} and c_{h2} per product respectively. The last term (3.7d) is the cost from expected returns of M periods ago, $p + c_r$ is the handling cost of a return plus the price of the product, as customer receive full return payment.

3.3.3 Bellman equation and value iteration

The objective of the MDP is to find the optimal policy that maximises the long-term discounted profit. The expected future discounted profit when following an optimal policy over n consecutive sub-periods when starting in state s is defined as $v_n(s)$. To obtain the optimal policy by value iteration one starts setting $v_0(s) = 0$ for all states s. Next, one computes for all $s : v_1(s) = \max_{a \in \mathcal{A}(I)} \{\mathbb{E}C_{\text{II}}(s,a)\}$, and continues by computing v_2, v_3, etc. using the recursive Bellman equation:

$$v_n(s) = \begin{cases} \max\limits_{q \in \mathcal{Q}(I)} \Bigg\{ \mathbb{E}C_{\text{I}}(s,q) + \max\limits_{a \in \mathcal{A}(I)} \Bigg\{ \mathbb{E}C_{\text{II}}(s,a) \\ \quad + \gamma \sum\limits_{d_1=0}^{D_1} \sum\limits_{d_2=0}^{D_2} \sum\limits_{m=1}^{M} \sum\limits_{r_m=0}^{R_m} P_1(d_1) P_2(d_2) B(r_m) \cdot v_{n-1}(s') \Bigg\} \Bigg\} & \text{if } n \bmod T = 0 \\ \max\limits_{a \in \mathcal{A}(I)} \Bigg\{ \mathbb{E}C_{\text{II}}(s,a) \\ \quad + \gamma \sum\limits_{d_1=0}^{D_1} \sum\limits_{d_2=0}^{D_2} \sum\limits_{m=1}^{M} \sum\limits_{r_m=0}^{R_m} P_1(d_1) P_2(d_2) B(r_m) \cdot v_{n-1}(s') \Bigg\} & \text{otherwise.} \end{cases} \tag{3.8}$$

Here, γ is the discount factor that is included to ensure a fair comparison between VI and the solution technique of DRL, which requires discounting of future rewards. The first equation in 3.8, where n is a multiple of T, relates to the ordering and rationing decision, which both happen at the beginning of the period consecutively. The second equation relates to the other time periods, in which only the rationing decision is involved.

Let n be the iteration counter. The Markov process is unichain and aperiodic at level I, and unichain but periodic at level II with period T. Hence the difference $v_n - v_{n-T}$ converges for all states to the same value, as n moves to infinity. A discounting factor γ is included for a fair comparison with DRL, which requires discounting for stabler training as future rewards might be uncorrelated in DRL. For $\gamma < 1$, the difference $v_n - v_{n-T}$ converges to zero, as n tends to infinity. If $\|v_n - v_{n-T}\|$ is smaller than $\varepsilon(1-\gamma)/\gamma$, the value iteration stops (we specify $\varepsilon = 0.1$). For details on the speed and conditions for convergence we refer to Puterman (1994). As at level II the problem is periodic with period T, one full iteration consists of T sub-periods.

The (nearly) optimal strategy for the two levels of the MDP can be obtained from the results of the value matrices. First the optimal replenishment policy $q^*(s)$ at $t = 1$ for level I is found by:

$$q^*(s) = \left\{ \begin{array}{l} \underset{q \in \mathcal{Q}(I)}{\arg\max} \left\{ \mathbb{E}C_{\mathrm{I}}(s,q) + \underset{a \in \mathcal{A}(I)}{\max} \left\{ \mathbb{E}C_{\mathrm{II}}(s,a) \right. \right. \\ \left. \left. + \gamma \sum_{d_1=0}^{D_1} \sum_{d_2=0}^{D_2} \sum_{m=1}^{M} \sum_{r_m=0}^{R_m} P_1(d_1) P_2(d_2) B(r_m) \cdot v_{n-1}(s') \right\} \right\} \end{array} \right. \tag{3.9}$$

Second, the optimal rationing decision $a^*(s)$ for all states in $t = 1, ..., T$ for level II is found by:

$$a^*(s) = \left\{ \begin{array}{l} \underset{a \in \mathcal{A}(I)}{\arg\max} \left\{ \mathbb{E}C_{\mathrm{II}}(s,a) \right. \\ \left. + \gamma \sum_{d_1=0}^{D_1} \sum_{d_2=0}^{D_2} \sum_{m=1}^{M} \sum_{r_m=0}^{R_m} P_1(d_1) P_2(d_2) B(r_m) \cdot v_{n-1}(s') \right\} \end{array} \right. \tag{3.10}$$

3.4 Implementation of Deep Reinforcement Learning Algorithm: Proximal Policy Optimisation

For small-scale settings the Bellman equations can be solved to optimality within reasonable computation time and RAM usage. However, for large-scale settings, the action and state space becomes increasingly large resulting in infeasible computational time or memory limitations. DRL is considered to be useful to circumvent the curse of dimensionality, as it can be used to develop near-optimal solutions that could not be obtained using conventional approaches (Boute et al., 2022). In DRL, many different algorithms are used. They all follow the general structure of an agent interacting with an environment and collecting experience to find the best policy. This policy is found by using the collected experience to train an approximation of the state values and policy, in which the approximations often take the form of a neural network.

Actor-critic methods are reinforcement learning algorithms in which both policy gradient and value estimation are applied. The actor and the critic are the two interrelated components making up the method. The critic provides an estimation of the expected future discounted profit, which is used by the actor to apply gradient descent on the policy. Asynchronous Advantage Actor-Critic (A3C) is a well-known actor-critic method that has proven to perform well for various inventory management problems (Gijsbrechts et al., 2022). However, the A3C algorithm has the disadvantage of needing a lot of training data and extensive parameter tuning. In order for the actor to converge, it needs to have an almost infinite amount of training data. Recently, Vanvuchelen et al. (2020) have shown that another actor-critic method called Proximal Policy Optimisation (PPO) also performs well in an inventory management setting and that it is less sensitive to the disadvantages of A3C. PPO is praised for its sample efficiency and ease of implementation (Schulman et al., 2017).

The PPO algorithm trains the actor and the critic, which are both represented by a neural network. The input of these neural networks is the state of the system. The output for the actor is the action and the output for the critic is an estimation of the expected future discounted profits for the input state. As the hierarchical MDP discussed in this paper has two levels, we use two actor-critic sets and four neural networks in this paper. The advantage of having a separate actor-critic set for each action is that it decreases the complexity of training the actor neural network. This could improve training, as the neural networks only have to approximate one action. The architecture of the neural networks is identical for all four networks, except for the output layer.

The neural networks consists of two fully connected layers with width 256, all with a tanh activation function and a bias. The tanh activation function has the advantage of non-linearity and has been proven to be better at quickly finding local (or global) minima as the derivatives can be large (LeCun et al., 2012). The four neural networks have different applications, and therefore the parameters of the neural network, which are the weights and bias, are different. The parameters of the actor and critic neural network are given by θ_h and ϕ_h respectively, where the level of the MDP is given by $h \in \{1 = \text{level I}, 2 = \text{level II}\}$.

For the actor, the last layer is a softmax activation function, which provides a probability distribution over the action space denoted by $\pi(\cdot|S, \theta)$ with S being the input state. Therefore, the size of the last layer is equal to the action space of the level. From the actor we can get the best action by taking the action with highest probability. For the critic, the size of the last layer is one, as the output is the estimation of the expected future profit for state S is denoted by $V(S, \phi)$.

3.4.1 PPO algorithm

To update the neural networks, the PPO algorithm needs to collect information about the environment. This is performed by sampling the studied problem and storing the visited states, actions, and related profits during the sampling. Figure 3.2 visualises the implementation of the PPO algorithm with the sampling, updating of the neural network, and the evaluation of the policy found. We sample for B periods and thus $T \cdot B$ sub-periods, and store the visited states, taken actions, and received profits. As our MDP has two levels, the storing of the training data differs and we store them separately. The algorithm starts with initialising the neural network parameters randomly and setting the training iteration counter (κ) to one.

3.4.2 Sampling

The sampling is started by drawing a random state S_1, setting the sub-period $t = 1$, and setting the sample points K and k to one. The state S_K is used in the actor and critic of level I to get the ordering action q_K and expected future profit $V(S_K, \phi_1)$. The ordering action q_K is obtained by randomly sampling from the probability distribution of the ordering action space, $q_K \sim \pi(\cdot|S_K, \theta_1)$. The ordering action is put into environment level I to get the state s_k, which is used by the actor and critic of level II. Additionally, from environment level I we collect the profit E_K when taking action q_K when in state S_k, this information combined with $V(S_K, \phi_1)$ is stored in the training data of level I. From the actor and critic of level II a rationing action is drawn from the probability distribution $a_k \sim \pi(\cdot|s_k, \theta_2)$ and expected future profit $V(s_k, \phi_2)$ is gained,

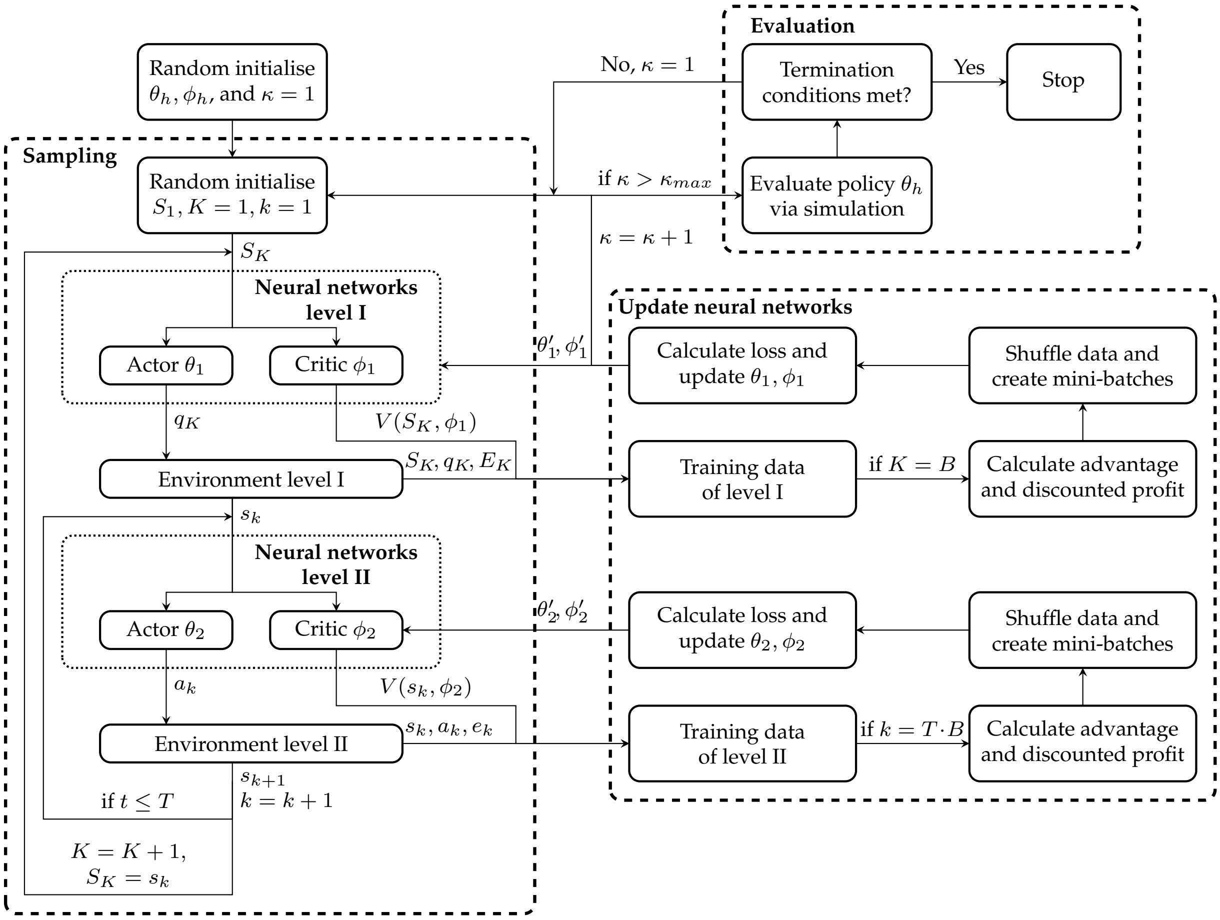


Figure 3.2: PPO algorithm flow diagram.

where the rationing action is put in environment level II to get the reward e_k, which together with the state and action is stored in training data of level II. The demand and returns that occur in the environment level II are randomly drawn from their respective distributions. Furthermore, the next state s_{k+1} is collected from environment level II. This next state is used as input for the neural network of level II if the sub-period has not ended yet, which is determined by $t \leq T$, else this state is used as input for the neural network of level I. Every time a sample is drawn from environment level II, both the sample point k and sub-period t are increased by one. When the sub-period has ended, the state s_k is used to construct state S_{K+1}, t is set to one indicating we are at the beginning of the period again, and the sample point K is increased by one. When $K = B$ and $k = T \cdot B$ the sampling is done and the neural networks are updated.

3.4.3 Update neural networks

First, with the collected training data the advantage and discounted profit are calculated. Secondly, the training data, advantage, and discounted profit are shuffled and split into η mini-batches to calculate the average loss over each mini-batch. Each mini-batch consecutively updates the neural networks once. For each mini-batch the loss is minimised by updating the neural networks parameters θ_h and ϕ_h to θ'_h and ϕ'_h. The procedure of calculating the loss and how the parameters of the neural network are updated with respective to the loss is given in Appendix 3.A.

The process of splitting the training data, advantage, and discounted profit and updating the neural networks is called an epoch. As we are using the collected training data, advantage, and discounted profits multiple times to update the neural networks a total of u epochs are performed. In addition, in every epoch the training data is shuffled, to create new configuration of mini-batches in each epoch. This ensures that the updates of the neural networks are not over-fitted to the training data. Thus, the parameters of the neural networks are updated $\eta \cdot u$ times, were each sample point in the training data is used u times.

3.4.4 Evaluation

Once the neural networks are trained u epochs, the training iteration counter κ is updated by one. After updating the neural networks for κ_{max} iterations, where each training iteration consists of B periods, the resulting policy is evaluated. We follow a similar procedure as Vanvuchelen et al. (2020). The policy is simulated for 1,000,000 periods considering a small warm-up period. During the sampling the action to be taken is randomly drawn from the output of the actor, which is a probability distribution over the actions, whereas in the evaluation we take the action with the highest probability as this is considered to be the best found action for the given state. Thus for the ordering action we take $q = \arg\max\left(\pi\left(\cdot|S, \theta_1\right)\right)$ and for the rationing action $a = \arg\max\left(\pi\left(\cdot|s, \theta_2\right)\right)$.

To measure the performance, the average profit per period from the simulation is used. If the average profit per period of the best resulting policy has not changed more than 0.5% over three consecutive policy evaluations, we assume the algorithm has converged to its final policy. A maximum of 30 policy evaluations are considered, if the maximum is reached the best performing policy is assumed to be the final policy.

Table 3.1: Parameter values used in the implementation of the PPO algorithm.

Hyperparameter	Symbol	Value
Depth of NN		2
Width of NN		256
Activation function		TanH
Sample periods	B	512
Amount of mini-batches	η	4
Number of epochs	μ	10
Training iterations	κ_{max}	1,000
Learning rate	α	10^{-4}
Entropy regularisation	β_E	10^{-5}
Clipping parameter	ϵ	0.2
Huber loss constant	δ	5

3.4.5 Implementation

Parameters

To implement the PPO algorithm, several parameters and input data need to be defined. The tuning of reinforcement learning parameters can be computational costly (Gijsbrechts et al., 2022), thus we started from the settings of related work such as Vanvuchelen et al. (2020). We did however adapt some of the parameters to our problem setting, and performed some experiments to improve the performance of our PPO algorithm. This was done by tuning the parameters and investigating the results for a test case. We evaluated different neural network architectures, sample periods, amount of mini-batches, learning rate, entropy regularisation, and the Huber loss constant. Table 3.1 describes the parameters used.

Our model has relatively large demand fluctuations during the period compared to related work, a larger sampling period is used to reduce uncertainty in the update of the network. Furthermore, the width of the neural network is increased, as our action space is relatively large and should not exceed the width of the neural network. The initial bias and weights of the neural network are sampled from a Normal distribution with mean 0 and standard deviation 0.5. Therefore the initial policy is a random policy.

Scaling of states and input data

By normalising the input data of the neural network it is expected to converge quicker to the training data (LeCun et al., 2012). The input of the neural network, which is the state, is therefore scaled so that the minimum and largest value of the state are -5 and 5. Additionally, the economic parameters are scaled from 0 to 1.

Action masking

As described in Section 3.3.2, the action space is limited by the current state for the replenishment and rationing action. To ensure the PPO algorithm does not consider invalid actions, we apply action masking using the method of Huang & Ontañón (2020). This method sets the output of invalid actions in the actor neural network before the softmax activation layer to a small negative value (in our case -10^{-8}), so that after the softmax activation layer the probability of choosing these actions becomes negligible. An additional advantage of action masking is that by masking the action space, the PPO algorithm can learn faster as it quickly learns to disregard the invalid actions.

3.5 Computational Complexity and Performance

As mentioned in Section 3.4, only small-scale instances can normally be solved with the Bellman equation. To compare the computational complexity and performance of PPO with VI, we develop a range of small-scale instances, consisting of a base test case and various alterations. The data set used is based on recent literature studying similar omnichannel retail settings with a few alterations to reduce the size of the problem (Li et al., 2015; Dijkstra et al., 2019; Ovezmyradov & Kurata, 2019; Bayram & Cesaret, 2021; Goedhart et al., 2022). Additionally, we investigate the effect of modelling the returns multi-period and sales-dependent by comparing it with two other methods that approximate the return flow.

3.5.1 Base test case and alterations

For the base test case, we assume average demand of the individual channels is assumed to be Poisson distributed with $\mu_1 = 3$ and $\mu_2 = 1$. As one needs a finite support, the distribution is right truncated at a cumulative probability of 99%, preserving the mean value using the approach by Cohen (1954) is used. Thus the maximum demand in a sub-period is $D_1 = 8$ and $D_2 = 4$.

We assume a period consists of 7 sub-periods ($T = 7$), to reflect a weekly ordering decision. Replenishment orders are placed on Monday ($t = 1$) with lead time $L = 2$,

Table 3.2: Parameter values for small-scale instances, where the base case is given in bold.

Parameter	Value	Value
Offline demand	μ_1	$1, 2, \mathbf{3}$
Online demand	μ_2	$\mathbf{1}, F(\rho)$
Coefficient of variation	CV_i	$1/(2\sqrt{\mu_i}), \boldsymbol{\sqrt{1/\mu_i}}$
Lead time	L	$1, \mathbf{2}, 3$
Return rate	ρ_p	$0.2, 0.3, \mathbf{0.4}$
Return window	M	$1, \mathbf{2}$
Procurement cost	c_p	$\mathbf{30}, 50, 70$
Return handling cost	c_r	$0, \mathbf{5}, 10$
Offline holding cost	c_{h1}	$0.1, 0.2, \mathbf{1}$
Online holding cost	c_{h2}	$0.05, 0.1, \mathbf{0.5}$

thus delivered on Wednesday ($t = 3$). Here both events occur at the beginning of the sub-period.

The probability of a product being returned is $\rho_p = 0.4$ in a return window $M = 2$ periods (Dijkstra et al., 2019). The probability of a product being returned in a sub-period is $\rho = 1-(1-\rho_p)^{1/(T\cdot M)} = 0.036$. Similar to the demand distribution, we truncate the return distribution at a cumulative return probability of 99%. Additionally, we reshape the binomial distribution according to the approach of Johnson et al. (2005) to preserve the mean value. Although a binomial distribution has a natural maximum value, we truncate the return distribution to reduce the size of the problem.

The economic parameters are based on literature studying similar settings (Li et al., 2015; Ovezmyradov & Kurata, 2019; Bayram & Cesaret, 2021). We set the price of the product at $p = 100$ and the cost of the product $c_p = 30$. The handling cost incurred for satisfying an online order is set at $c_u = 5$. The holding cost of the offline and online channel are $c_{h1} = 1$ and $c_{h2} = 0.5$, respectively. The handling cost of a returned product is $c_r = 5$. The discount factor is $\gamma = 0.99$, similarly as in Vanvuchelen et al. (2020).

The different instances are created by varying specific subsets of parameters. Table 3.2 presents the values for the specific parameters. The parameter values for the base case are listed in bold.

The different instances in which $\mu_2 = F(\rho)$ represent scenarios in which the online return percentages vary, but the net online sales remains constant. These instances are created to investigate how the uncertainty of returns influence the profitability of the retailer, without it being influenced by the net online sales. The online demand is calculated as follows:

$$F(\rho) = \frac{\mu_2^*(1-\rho_p^*)}{(1-\rho)} \tag{3.11}$$

Here, μ_2^* and ρ_p^* are the online demand and return percentage of the base test case. Other instances are created by halving the coefficient of variation (CV_i). For alternative values of CV_i, the demand distribution is fitted to μ_i and CV_i using the procedure described in Adan et al. (1995). All demand distributions are reshaped as right-truncated distributions (Cohen, 1954; Shah, 1966; Johnson et al., 2005; Louchard & Ward, 2015), truncating at a probability of 99% of the cumulative distribution function. In the remainder of this paper, if a instance does not specify a certain parameter value, it will be equal to their base test case setting.

3.5.2 Value Iteration vs Proximal Policy Optimisation

The results presented in this section were obtained by implementing VI in Python version 3.7.2. For the neural network models of the PPO algorithm TensorFlow 2.3.0 was used. The model was run on a Personal Computer with Intel Xeon W-2133 CPU @ 3.60 GHz and 32GB of RAM.

Table 3.3 presents the required computational time, RAM usage, and optimality gap for both algorithms, as well as the total amount of states in the system. The RAM usage is mainly dependent on the size of the value matrix (as this changes due to the ordering state becoming zero after replenishment, it is expressed as the maximum RAM usage during an iteration). The optimality gap is the difference in profit between the optimal solution derived via VI and the policy found by the PPO algorithm. The profit for the policy found by VI and PPO is obtained by simulating the resulting policies of both algorithms for 1,000,000 periods with a small warm-up period.

From Table 3.3, it is observed that the demand, coefficient of variation, returns, and lead time increase the state space exponentially. Especially for the return percentages the curse of dimensionality is clear, as an increase of 10% in return percentages doubles the amount of states in the model. For some return percentages, a similar amount of states can be seen. This is due to the truncation of the distribution function for the number of returns in a sub-period, as the truncation then happens at the same value. Increasing the lead time increases the number of sub-periods the ordering state

Table 3.3: Computational complexity and performance

for VI and the PPO algorithm for the small-scale instances.

Instance	States (in millions)	CPU time (hr) MDP	CPU time (hr) PPO	RAM (MB) MDP	RAM (MB) PPO	Optimality gap (%)
$\mu_1 = 3$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$\mu_1 = 2$	153.6	12.19	104.2	3170.0	5.4	−2.20
$\mu_1 = 1$	108.8	7.61	146.6†	2219.5	4.7	−2.73
$CV_i = 1/\sqrt{\mu_i}$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$CV_1 = 1/(2\sqrt{\mu_1})$	130.3	6.04	141.8	2674.5	5.1	−1.72
$CV_2 = 1/(2\sqrt{\mu_2})$	59.3	0.55	142.3	1275.8	5.0	−1.84
$CV_i = 1/(2\sqrt{\mu_i})$	33.7	0.24	142.0	709.1	3.9	−1.33
$\rho_p = 0.2$	43.8	12.72	128.8	521.9	5.3	−2.12
$\rho_p = 0.3$	107.1	17.48	100.3	1746.2	5.7	−2.29
$\rho_p = 0.4$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$\rho_p = 0.5$	431.8	30.24	145.8†	8777.4	6.5	−2.20
$\rho_p = 0.6$	431.8	31.93	146.1†	8777.4	6.5	−2.03
$\rho_p = 0.2, \mu_2 = F(\rho_p)$	18.8	12.72	123.3	440.4	5.0	−2.12
$\rho_p = 0.3, \mu_2 = F(\rho_p)$	80.3	17.48	149.2†	1741.2	5.7	−2.29
$\rho_p = 0.4, \mu_2 = F(\rho_p)$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$\rho_p = 0.5, \mu_2 = F(\rho_p)$	431.8	30.24	123.3	8777.4	5.0	−2.20
$\rho_p = 0.6, \mu_2 = F(\rho_p)$	903.4	31.93	123.3	18131.9	5.0	−2.03
$M = 1$	10.5	0.85	147.0†	149.6	5.7	−2.33
$M = 2$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$c_r = 0$	205.8	22.63	145.6†	4283.1	6.1	−2.11
$c_r = 5$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$c_r = 10$	205.8	22.78	148.6†	4283.1	6.1	−2.44
$c_p = 30$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$c_p = 50$	205.8	23.41	144.4†	4283.1	6.1	−2.04
$c_p = 70$	205.8	24.33	126.4	4283.1	6.1	−0.89
$c_{h1} = 0.1, c_{h2} = 0.05$	205.8	23.96	125.0	4283.1	6.1	−1.97
$c_{h1} = 0.2, c_{h2} = 0.1$	205.8	24.40	126.9	4283.1	6.1	−0.81
$c_{h1} = 1, c_{h2} = 0.5$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$L = 1$	63.3	3.30	135.9	3325.7	5.4	−2.40
$L = 2$*	205.8	17.87	85.7	4283.1	6.1	−1.64
$L = 3$	493.3	58.63	137.9	5360.8	6.9	−2.34
Average	214.9	18.39	126.9	4251.3	5.6	−1.95

*Base test case with $\mu_1 = 3, \mu_2 = 1, CV_i = 1/\sqrt{\mu_i}, \rho_p = 0.4, M = 2, c_r = 5, c_p = 30, c_{h1} = 1, c_{h2} = 0.5, L = 2$

† Instances in which the policy did not converge within 30 iterations

is included in the state space and the maximum order quantity, therefore increasing the amount of states significantly.

It is observed that for the small-scale instances the CPU time of VI is much lower compared to the PPO algorithm. The CPU time of VI grows exponentially with the number of states and state transitions, the CPU time of the PPO algorithm is less sensitive to the dimensionality of the instance. VI solves the problem iteratively backwards via the Bellman equation, where for each possible state, the action and related probability distribution on the next states is calculated and used to update the state value. Therefore, the CPU time of VI does not solely depend on the number of states but also on the number of state transitions and the action space resulting in larger computational time.

The PPO algorithm is less sensitive to the dimensionality of the problem as it approximates the state values and actions by a neural network. The results of the simulation are used to update the neural network, which generalises across similar states. As the PPO algorithm does not need to visit each state to solve the problem, the CPU time does not increase with the number of states. However, as the PPO algorithm is less effective in finding the optimal action for each state compared to VI, the computational time is often larger. The PPO algorithm uses simulation and random initial weights of the neural network, varying the computational time used for training per instance. Furthermore, the visited states, rewards, and actions are randomly simulated, influencing the training as this information is used to update the neural network.

From Table 3.3 it can be concluded that the RAM usage of the MDP increases with the size of the instance. The RAM usage of the MDP is less sensitive to the demand and lead time, as these increase the size of the value matrix to a limited extent. The RAM usage is more sensitive to the return distribution, as this influences the state space of multiple states. The RAM usage of the PPO algorithm is less sensitive to parameter values as the only information it needs to store are the parameters of the neural network and the samples of the simulation. These are less dependent of the size of the problem.

From Table 3.3 it can be observed that the largest optimality gap is found for the instance of $\mu_1 = 1, \mu_2 = 1$. Here, the PPO algorithm yields 2.73% less profit than the optimal solution. The PPO algorithm performs best for the instance of $c_{h1} = 0.2, c_{h2} = 0.1$, in which the profit gap is -0.81%.

When looking at the underlying cost and how the found policy of the PPO algorithm differs from the optimal policy (see also Table 3.7 in Appendix 3.B and Table 3.8 in Appendix 3.C), it is observed that instances in which the PPO algorithm has difficulty in finding the optimal policy is when around the optimal solution the profit difference is minimal. Finding the optimal action is then difficult as performing non-optimal minimally influences the profit. This is observed as the rationing decision at high inventory level differs the most between the two policies. At a high inventory level the rationing decision has less influence on the profit as there is no competition between the channels for products. Therefore, at high inventory levels the rationing decision is more focused on cost minimisation. Additionally, when the action state space is large the PPO algorithm has more decisions to evaluate, increasing the difficulty of finding the optimal action. Consequently, at low inventory levels, the PPO algorithm can more easily identify the optimal rationing decision as the action space is limited by the inventory level and the channels compete for products, thus the rationing decision also influences the profit more.

For the instances with decreasing demands, we notice that the optimality gap is growing. A similar trend can be observed for increasing coefficients of variation in demand. This suggests that having a higher dispersion of the demand distribution negatively influences the training of the PPO algorithm. With a higher coefficient of variation, the rewards that the PPO algorithm experiences during sampling are more dispersed as this is partly driven by demand. To better distinguish good actions from bad actions, more training data is needed to mitigate the influence of the larger dispersed demand distribution.

Overall, the PPO algorithm shows good performance, with a profit optimality gap of around -2.0%. The advantage of the PPO algorithm is that it is not influenced much by the dimensions of the studied problem, as CPU time and RAM usage are relatively stable. The PPO algorithm does however experience difficulties in finding the optimal policy when the uncertainty increases. Furthermore, the rewards influence the training, as these influence the trade-off the PPO algorithm makes between different actions. As the training of a reinforcement learning algorithm is influenced by the received rewards, and one can manipulate them to find better policies, this is often referred to as reward shaping in the reinforcement learning literature.

3.5.3 Modelling returns

Current literature mentions several methods to model returns. In this paper, we explicitly model the returns as multi-period and sales-dependent (which we refer to as Multi R in Table 3.4). To investigate how much influence this approach has on the retailer's inventory management, we compare it with two existing modelling approaches: (1) sales-aggregated returns (Agg. R) and (2) sales-independent returns (Ind. R).

For the approach of Agg. R, we model the returns by aggregating the state $\boldsymbol{R}$ into a single state R and omit R_0. The state R now approximates the total online sales in the past M periods. However, as the expiration date of whether a products can be returned is not included in the state it needs to be approximated. We therefore define the state transition of R to R' as follows:

$$R' = (R - r - h + \min(d_2, I - a))^+ \tag{3.12}$$

Here, r represents the number of products being returned (determined by a Binomial distribution with ρ as the probability a single product is returned), R is the pool of returnable items, h is the approximated online sales that cannot be returned anymore, and the last term is the online sales of the current sub-period. We approximate h using a Poisson distribution with an average value of $\mu_2 \cdot (1 - \rho_p)$ as this the number of products that on average are not returned.

Modelling the returns independent of sales, as we do in Ind. R, results in returns being an exogenous flow, similar to having a negative demand (see e.g. Feinberg & Lewis, 2005). We approximate the return flow as a negative demand with a Binomial distribution with ρ as the probability a single product is returned, and average online sales over the return window $(\mu_2 \cdot M \cdot T)$ as the pool of returnable items. With regards to our MDP, the states R_0 and $\boldsymbol{R}$ are not relevant anymore when approximating the returns as independent of sales and are therefore omitted.

We reformulate the MDP for both methods of modelling returns and use VI to obtain the policies. Similar to Section 3.5.2, we compare the different methods of modelling returns on the optimality gap, the required computational time, and the RAM usage. The profit is calculated by simulation the resulting policies in the MDP with the multi-period sales-dependent returns to investigate the optimality gap. The results are presented in Table 3.4.

The advantage of modelling the returns independent from demand (or past sales) is clearly seen in the CPU time. Modelling the returns based on an aggregated number seems to be more influenced by the increased state spaces, but the CPU time still remains low compared to the original MDP. Similar trends can be observed for the RAM

Table 3.4: Computational complexity and performance for the different modelling approaches.

Instance	CPU time (hr)			RAM usage (MB)			Optimality gap (%)	
	Multi R	Agg. R	Ind. R	Multi R	Agg. R	Ind. R	Agg. R	Ind. R
$\mu_1 = 3$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$\mu_1 = 2$	12.19	1.13	0.37	3170.0	124.1	31.0	−1.57	−14.87
$\mu_1 = 1$	7.61	0.58	0.85	2219.5	69.6	15.9	−3.04	−33.44
$CV_i = 1/(\sqrt{\mu_i})$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$CV_1 = 1/(2\sqrt{\mu_1})$	6.04	0.61	0.20	2674.5	94.5	26.6	−1.11	−15.56
$CV_2 = 1/(2\sqrt{\mu_2})$	0.55	0.72	0.17	1275.8	127.3	22.2	−0.97	−3.23
$CV_i = 1/(2\sqrt{\mu_i})$	0.24	0.31	0.13	709.1	57.4	10.7	−1.04	−6.38
$\rho_p = 0.2$	12.72	0.49	0.17	521.9	73.7	42.1	−0.03	−2.99
$\rho_p = 0.3$	17.48	1.31	0.20	1746.2	129.2	42.1	−0.88	−5.00
$\rho_p = 0.4$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$\rho_p = 0.5$	30.24	2.64	0.33	8777.4	464.6	70.7	−0.78	−19.27
$\rho_p = 0.6$	31.93	2.71	0.52	8777.4	613.1	78.2	−0.37	−36.93
$\rho_p = 0.2, \mu_2 = F(\rho_p)$	2.59	0.40	0.15	440.4	40.5	12.8	−0.17	−3.12
$\rho_p = 0.3, \mu_2 = F(\rho_p)$	17.74	1.25	0.20	1741.2	105.5	38.0	−0.74	−6.79
$\rho_p = 0.4, \mu_2 = F(\rho_p)$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$\rho_p = 0.5, \mu_2 = F(\rho_p)$	28.70	3.11	0.30	8777.4	464.6	70.7	−1.00	−17.07
$\rho_p = 0.6, \mu_2 = F(\rho_p)$	83.63	5.40	0.78	18131.9	992.8	148.5	−0.73	−57.78
$M = 1$	0.85	0.50	0.23	149.5	73.7	27.3	−1.40	−10.73
$M = 2$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$c_r = 0$	22.63	1.55	0.23	4283.1	199.7	53.0	−0.76	−8.83
$c_r = 5$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$c_r = 10$	22.78	1.61	0.23	4283.1	199.7	53.0	−1.39	−8.97
$c_p = 30$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$c_p = 50$	23.41	2.19	0.25	4283.1	199.7	53.0	−1.78	−14.55
$c_p = 70$	24.33	1.96	0.25	4283.1	199.7	53.0	−2.51	−29.22
$c_{h1} = 0.1, c_{h2} = 0.05$	24.40	1.83	0.23	4283.1	199.7	53.0	−0.80	−5.69
$c_{h1} = 0.2, c_{h2} = 0.1$	23.96	1.86	0.23	4283.1	199.7	53.0	−0.83	−6.24
$c_{h1} = 1, c_{h2} = 0.5$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$L = 1$	3.30	0.82	0.22	3325.7	152.2	37.6	−0.97	−7.20
$L = 2$*	17.87	1.82	0.23	4283.1	199.7	53.0	−1.03	−8.88
$L = 3$	58.63	3.59	0.27	5360.8	253.6	71.0	−1.04	−12.40
Average	20.60	1.67	0.29	4251.32	227.58	48.54	−1.08	−14.57

*Base test case with: $\mu_1 = 3, \mu_2 = 1, CV_i = 1/\sqrt{\mu_i}, \rho_p = 0.4, M = 2, c_r = 5, c_p = 30, c_{h1} = 1, c_{h2} = 0.5, L = 2$

usage. It can be concluded that there is a clear trade-off: exact modelling results in a better performing policy at the cost of computational complexity.

From Table 3.4, it is observed that the sales-aggregated return method shows a low optimality gap, between -3.04% and -0.03%. When online sales become more dominant a larger optimality gap is observed and with lower return percentages the optimality gap decreases. The sales-independent return method shows larger optimality gaps, as high as -60%. Here, the optimality gap increases when the returns are becoming more dominant, such as in scenarios with larger shares of online sales or higher return percentages. Additionally, for both methods the optimality gap increases with higher cost of the product.

Overall, the approach of aggregating the sales still shows relatively good performance, as the optimality gap is lower compared to the policies found by the PPO algorithm. Both methods show good performance and have the potential to be used for large-scale instances. We further elaborate on the performance of the aggregated approach and the PPO algorithm for large-scale instances in the next section.

3.6 Large-scale Instances

In many settings in the literature and in practice, demand volumes, lead times, return windows, and return percentages are larger than in the small-scale settings discussed above. In such settings, the state space of the MDP gets too large to obtain an exact solution. For instance, for the presented large-scale base case, the amount of states would be just over one billion, with a RAM usage of around 16 GB, and an estimated CPU time of 300 hrs if solved with VI. Therefore, we use the PPO algorithm with the multi-period sales-dependent returns and the MDP in which the sales are aggregated to solve and analyse large-scale instances. From these large-scale instances relevant managerial insights can be derived as they better reflect current practices. The cost parameters and coefficient of variation of the small-scale instances are also used for the large-scale instances, but parameters that influence the state space of the MDP are increased. In Table 3.5 the different parameter values for the large-scale instances are presented.

Table 3.5: Parameter values for large-scale instances, where the base case is given in bold.

Parameter	Symbol	Value
Offline demand	μ_1	$2, 4, \mathbf{6}$
Online demand	μ_2	$\mathbf{2}, 4, 6, F(\rho_p)$
Return rate	ρ_p	$0.3, \mathbf{0.4}, 0.5, 0.6$
Return window	M	$\mathbf{2}, 3, 4, 5$
Lead time	L	$1, \mathbf{2}, 3, 4$

We evaluate the different instances on different performance indicators related to profit and service level. We focus on instances in which demand, return percentages and windows, and lead time are altered as these instances give us insight in how uncertainty influences omni-channel retailer's performance. We investigate the profit and the alpha service level, which is an important performance indicator that tells us the fraction of time one ends a day ends with products still in stock. Most important is the cycle service level (i.e. the alpha service level just before replenishment). Table 3.6 gives the profit and cycle service level for our large-scale instances. These results were obtained by simulating the policy found by the PPO algorithm and from VI of the sales-aggregated return MDP with the same procedure as described in Section 3.5.2.

For almost all the small-scale instances presented in Section 3.5, the policy resulting from the sales-aggregated return MDP showed a lower optimality gap than the policy of the PPO algorithm. Table 3.6 shows that for large-scale instances, however, the policy of the PPO algorithm has a higher profit than the policy of the sales-aggregated return MDP. The cycle service level is higher for the policy of the PPO algorithm across almost all instances and channels, indicating that more demand is fulfilled.

From Table 3.6 it is also observed that if the average demand shifts from the offline channel to the online channel, the profit decreases. This is mostly the result of the increase in return cost and lower net sales: if more products are ordered online, the return flow will be higher. Although the profit decreases with more online sales, the policies found differ on the cycle service level. Both policies have higher cycle service level for the offline channel when offline sales decrease. However, where the policy of the PPO algorithm decreases the service level for the online channel with higher online sales, the policy of the sales-aggregated return MDP increases their service level. Overall, the policy of the PPO algorithm has a higher service level across all instances.

The return percentage negatively influences the profit, as more products are returned the retailer has higher return costs. Although the retailer orders less products in instances with higher return percentages, the cycle service level is one of the highest for instances with the highest return percentage. As the retailer has less influence on their inventory position, they might have excessive stock before replenishment arrives. For the instances in which the net online sales remains constant, an opposite trend is observed for the online channel. As the return percentage and online demand increases, the cycle service level decreases except for the offline channel with the policy found by

Table 3.6: Profit and cycle service level for large-scale instances.

	Profit		Cycle service level			
			Offline		Online	
Instance	PPO	Agg. R	PPO	Agg. R	PPO	Agg. R
$\mu_1 = 6, \mu_2 = 2$*	3130.28	3061.15	0.982	0.969	0.950	0.806
$\mu_1 = 4, \mu_2 = 4$	2662.31	2596.45	0.987	0.986	0.945	0.832
$\mu_1 = 2, \mu_2 = 6$	2226.48	2131.20	0.990	0.986	0.936	0.845
$\rho_p = 0.3$	3241.45	3170.67	0.979	0.970	0.945	0.852
$\rho_p = 0.4$*	3130.28	3061.15	0.982	0.969	0.950	0.806
$\rho_p = 0.5$	3033.68	2967.38	0.988	0.970	0.955	0.800
$\rho_p = 0.6$	2934.70	2877.55	0.993	0.974	0.958	0.825
$\rho_p = 0.3, \mu_2 = F(\rho_p)$	3102.50	3081.22	0.978	0.967	0.945	0.833
$\rho_p = 0.4, \mu_2 = F(\rho_p)$*	3130.28	3061.15	0.982	0.969	0.950	0.806
$\rho_p = 0.5, \mu_2 = F(\rho_p)$	3081.73	3038.53	0.982	0.973	0.945	0.796
$\rho_p = 0.6, \mu_2 = F(\rho_p)$	3026.83	3004.43	0.977	0.980	0.903	0.812
$M = 2$*	3130.28	3061.15	0.982	0.969	0.950	0.806
$M = 3$	3134.32	3048.15	0.986	0.968	0.952	0.829
$M = 4$	3136.88	3042.26	0.968	0.965	0.959	0.850
$M = 5$	3138.79	3030.63	0.991	0.966	0.964	0.856
$L = 1$	3137.64	3065.18	0.981	0.969	0.953	0.806
$L = 2$*	3130.28	3061.15	0.982	0.969	0.950	0.806
$L = 3$	3127.76	3058.87	0.963	0.968	0.946	0.800
$L = 4$	3112.90	3055.85	0.944	0.968	0.907	0.797
Average	3015.22	2948.64	0.979	0.972	0.944	0.823

*Base test case with: $\mu_1 = 6$, $\mu_2 = 2$, $CV_i = 1/\sqrt{\mu_i}$, $\rho = 0.4$, $M = 2$, $c_r = 5$, $c_p = 30$, $c_{h1} = 1$, $c_{h2} = 0.5$, $L = 2$

the sales-aggregated return MDP. The profit decreases more in these instances, due to higher return flow, thus the retailer orders less. The retailer prefers to store products in the online channel as the holding cost is lower and make them available later during the sub-period for the offline channel if needed. However, as the return flow is too high, storing them in the online channel becomes less profitable as they might be sold to online customers. Therefore, the retailer orders less products for all channels and will face a stock-out more frequently. For the policy found by the sales-aggregated return MDP this also happens, however the rationing action differs where more products are stored in the offline channel with increasing online sales.

When investigating longer return windows it is observed that the profit is not much influenced. Although the inventory cost is slightly increased, they are negligible indicating that the return window has little effect on the profit for the retailer. However, the return window does influence the cycle service level. When the return window is increased, the cycle service level also increases. When the return window is increased, the probability of a product being returned in a sub-period is decreased. Therefore, the return distribution becomes less uncertain thus the retailer has more control on their inventory levels. Therefore they can better satisfy demand without ending up with excessive or little stock and thus improve their service level without high costs.

With an increase in lead time the profit does not change much, and for the policy of the sales-aggregated return MDP the cycle service level remains constant. However, for the policy found by the PPO algorithm the cycle service level does decrease. The policy found by the PPO algorithm has more difficulty with the uncertainty of demand during the lead time resulting in higher chance of a stock-out. However, it appears the cycle service level has little effect on the profit.

From Table 3.6 it can be concluded that the policy of the PPO algorithm outperforms the policy of the sales-aggregated return MDP for large-scale instances on both profit and cycle service level. Furthermore, a higher online demand and return flow negatively influences the profit. The cycle service level of the offline channel increases with higher return percentage, unless the online demand also increases as the retailer compensates for the decrease in profit margins by ordering less products thus negatively influencing the offline channel. However, the retailer can compensate for the lower inventory by preferring the offline channel via the rationing action. Furthermore, a longer return window or lead time has limited influence on the profit but does influence the cycle service level.

3.7 Conclusion and Discussion

Omni-channel retailers are experiencing an increase in returns originating from the growth of online sales. This return flow is becoming a significant issue for their inventory management because the uncertainty of whether a product is being returned can result in excessive stock. Also, several studies suggest to use the store for the fulfilment of online orders and the handling of returns. This has the advantage of leveraging the assets of the offline channel for the online channel.

In this paper, we therefore study the problem of a retailer who replenishes and rations their store inventory across an offline and online channel, as well as integrates returns in managing their inventory. The retailer has to make a trade-off between serving in-store and online customers, where the online sales might lead to returns. This paper contributes to the academic literature by providing a model for the inventory management of an omni-channel retailer with multi-period sales-dependent returns. In contrast to previous work, our model considers the returns dependent on sales over multiple periods. The resulting MDP can be solved with value iteration for small-scale instances resulting in exact solutions. However, the complexity of the model grows for larger demands and longer return windows and becomes unsolvable. Therefore, we investigate alternative solution methods such as DRL and approximating the returns.

As a DRL algorithm, PPO is used, as it has been proven to perform well in inventory management settings. The PPO algorithm is able to provide solutions within reasonable optimality gaps. Although VI shows a much lower running time for small instances, it increases exponential in the problem size. The PPO algorithm shows a high computation time for small instances but scales much better to larger problems. The PPO algorithm leaves an optimality gap of about 2.0% for all small-scale instances. Instances in which there is a higher level of uncertainty negatively influence the training of the PPO algorithm. Furthermore, the PPO algorithm has trouble finding the optimal action when the profit trade-off is minimal between the optimal and non-optimal actions. We conclude that the PPO algorithm is useful in environment with relatively low uncertainty and when the cost differences around the optimal solution are not too small.

When investigating different methods to model and approximate the return flow it was found that modelling the returns independently from historical sales showed a high optimality gap, increasing with larger return flows. Furthermore, aggregating historical sales in one state outperforms the PPO algorithm for small-scale instances (with an optimality gap of around 1.0%). Approximating the returns into one state has the advantage of lower CPU time and RAM usage, and in situations with relatively low return flows, the near-optimal results suggest that the existing models might be useful for implementations in practice.

As the multi-period sales-dependent returns MDP cannot be solved with VI for more realistic retail settings, we use PPO and VI on the MDP formulated with sales-aggregated returns to investigate how return and demand uncertainty influence the omni-channel retailer's profit and service level. For large-scale instances, the PPO algorithm outperforms the policy found with VI of the sales-aggregated return MDP, therefore the PPO algorithm is preferred for finding a policy when faced with larger demand volumes. The results indicate that if customers shift from the offline to the online channel it has a negative impact on the profit for the retailer. As online demand increases, the online returns increase which negatively influence the retailer profit as it requires more handling. If the online returns are becoming too large it will also negatively influence the service level of the offline channel, as the retailer will order less products for both channels due to the low profitability per product sold online unless the retailer compensate it with their rationing strategy. Furthermore, although a longer return window does not affect the profit per product sold, it does increase the service level of both channels, therefore a longer return window might be more preferable as long as it does not result in higher return percentages.

This paper investigates how returns influence the profitability of an omni-channel retailer who uses their store for the fulfilment of online demand and returns of online products. Although the model captures most typical characteristics of omni-channel retailers, several potentially relevant factors were not taken into account, and could be directions for further research. First, we only investigate the perspective of a single store. In a multi-store context, additional research could also include which stores are allowed to fulfil online demand and handle online returns. Additionally, we only consider online returns while some retailers might also experience a lot of returns from offline sales. It would be interesting what these offline sales have on the inventory management of the retailer. Furthermore, we have assumed that the return probability is constant over time. However, in practice retailers might experience that products that are sold recently are returned with a higher probability as the consumer immediately knows if they want to return it. Therefore, it is interesting to research different return probability distributions and their effect on the inventory management.

Second, we investigated different methods on modelling returns and their performance. However, an interesting research direction is to not aggregate the returns and use the PPO algorithm to find an optimal policy for this approach. Additionally, one might be able to circumvent the curse of dimensionality by better investigating if all states should be included in the MDP, as some might only be visited with such a low probability that they are not relevant for finding the optimal policy.

Third, further development of DRL algorithms for inventory problems is relevant. Although the context of our studied problem does not require discounting of future rewards, the PPO algorithm needs it for stable training and convergence. Having the discount factor too close to one results in unstable training, therefore an interesting research direction would be in the development of stable undiscounted DRL algorithms. The PPO algorithm develops a well-performing policy in our study, it has difficulties learning in some instances. Several research directions can be identified to improve the training of the PPO algorithm. Transfer learning (in which the policy found of one instance is used for the learning of similar instances) might be useful in reducing training time. Also, reward shaping might be used to provide better feedback on which actions are preferred. Furthermore, behavior cloning, in which the PPO algorithm is pre-trained with an expert policy such as a simple heuristic, might be useful. Further reductions in CPU time could also be achieved by parallelisation of the sampling of training data.

Appendix 3.A Loss function of PPO

Below we describe the procedure of calculating the loss for each set of samples in a mini-batch. The loss is used to update the parameters of the actor and critic neural network. The set of samples in a mini-batch for level I is denoted by $\mathcal{K}_1$ and for level II by $\mathcal{K}_2$, where $\mathcal{K}_1 = \{0, 1, \ldots, B/\eta\}$ and $\mathcal{K}_2 = \{0, 1, \ldots, T \cdot B/\eta\}$. A sample point at level I is denoted by K and at level II by k.

3.A.1 Advantage and discounted profit

The advantage and discounted profit are calculated for each sample in the mini-batch. The advantage is an indication of how well the chosen action performs compared with the expected cost. We use this information to update our neural networks. In this paper, we use the generalised advantage estimator approach, as described in Vanvuchelen et al. (2020). The advantage of sample K of level I is calculated as follows:

$$G_K = U_K - V\left(S_K, \phi_1\right) \tag{3.13}$$

Here, G_K is the advantage of sample K of level I and U_K is the discounted profit for sample K (defined below). The advantage of sample k of level II is calculated as follows:

$$g_k = u_k - V\left(s_k, \phi_2\right) \tag{3.14}$$

Here, g_k is the advantage of sample k of level II and u_k is the discounted profit for sample k (defined below). The discounted profit of sample K of level I is calculated as follow:

$$U_K = \sum_{i=K}^{B} \gamma^{i-K} \cdot E_i + \gamma^{B-K} \cdot V\left(S_B, \phi_1\right) \tag{3.15}$$

The discounted profit is an estimation of the future profit the retailer can expect to receive at it current state. This estimation is derived from the profit obtained through the sampling, and the value of the actor of the last state. The discount factor (γ) is needed to ensure the training of the PPO algorithm is stable and leads to convergence (Wiering, 2004). The discounted profit of sample k of level II is calculated similarly:

$$u_k = \sum_{i=k}^{T \cdot B} \gamma^{i-k} \cdot e_i + \gamma^{T \cdot B - k} \cdot V\left(s_{T \cdot B}, \phi_2\right) \tag{3.16}$$

3.A.2 Training the Neural Networks

With the collected training data, advantage, and discounted profit of the sampling, the gradient of the loss function with respect to the weights and bias of the neural networks can be calculated. The loss function is a predictor for the error of the neural network. By updating the weights and bias in the direction of the gradient of the loss with a step size, the loss can be minimised.

To train the neural network, three different type of loss functions are used, two for the actor network and one for the critic network. The loss functions of the actor network consist of the policy loss and an entropy loss. The policy loss trains the actor neural network so that actions that give high expected profit are preferred above actions with lower expected profits, while the entropy loss tries to encourage exploration of new actions. The policy loss for a shuffled mini-batch of level I is defined as follows:

$$\text{Policy loss level I} = -\sum_{K \in \mathcal{K}_1} \min\left(\frac{\pi(q_K|S_K, \theta_1')}{\pi(q_K|S_K, \theta_1)} \cdot G_K, \text{clip}\left(\frac{\pi(q_K|S_K, \theta_1')}{\pi(q_K|S_K, \theta_1)}, 1-\epsilon, 1+\epsilon\right) \cdot G_K\right) \tag{3.17}$$

Here, $\frac{\pi\left(q_K|S_K, \theta_1'\right)}{\pi(q_K|S_K, \theta_1)}$ is the ratio of the probability of choosing ordering action q_K in state S_K with the new neural network parameters θ_1' and the current parameters of the neural network θ_1. The policy loss formula uses a clipping formula to limit the loss function, where clip (x, x_{min}, x_{max}) ensures that x is between the range of x_{min} and x_{max}, otherwise the value is clipped to the range edges. The clipping parameter ϵ determines the value of the range edges.

The policy loss for a shuffled mini-batch of level II is defined as follows:

$$\text{Policy loss level II} = -\sum_{k \in \mathcal{K}_2} \min\left(\frac{\pi(a_k|s_k, \theta_2')}{\pi(a_k|s_k, \theta_2)} \cdot g_k, \text{clip}\left(\frac{\pi(a_k|s_k, \theta_2')}{\pi(a_k|s_k, \theta_2)}, 1-\epsilon, 1+\epsilon\right) \cdot g_k\right) \tag{3.18}$$

The entropy loss for a shuffled mini-batch of level I is defined as follows:

$$\text{Entropy loss level I} = \beta_E \sum_{K \in \mathcal{K}_1} \pi\left(\cdot|S_K, \theta_1'\right) \cdot \log \pi\left(\cdot|S_K, \theta_1'\right) \tag{3.19}$$

If $\pi(\cdot|S_K, \theta_1')$ is evenly distributed, thus each action has the same probability, the entropy will be largely negative. A deterministic policy, where one action has a high probability compared to others, results in the entropy loss to be closer to zero. The entropy regularisation term β_E determines how much emphasis is placed on the entropy in the loss when combined with the policy loss. By minimising the entropy, the probability distribution of actions is more evenly distributed. During sampling this encourages the algorithm to explore new actions, preventing the policy from converging to a bad-performing local optimal. The entropy loss for a shuffled mini-batch of level II is defined as follows:

$$\text{Entropy loss level II} = \beta_E \sum_{k \in \mathcal{K}_2} \pi\left(\cdot|s_k, \theta_2'\right) \cdot \log \pi\left(\cdot|s_k, \theta_2'\right) \tag{3.20}$$

For the critic network a value loss function is used, which is based on the difference between the future discounted profit and the output of the value function approximation. Here the Huber loss is used as it is less sensitive towards outliers as opposed to the mean squared error loss (Huber, 1964). Due to potentially large demand fluctuations, such outliers in profit are quite likely to ocur in our problem setting, and the use of the Huber loss therefore is less influenced by these outliers. The Huber loss is defined as follows:

$$L(x) = \begin{cases} \frac{1}{2}x^2 & \text{for } |x| \leq \delta \\ \delta\left(|x| - \frac{1}{2}\delta\right) & \text{otherwise.} \end{cases} \tag{3.21}$$

Where δ is the Huber loss constant. The value loss for a shuffled mini-batch of level I is defined as follows:

$$\text{Value loss level I} = \sum_{K \in \mathcal{K}_1} L\left(V\left(S_K, \phi_1'\right) - U_K\right) \tag{3.22}$$

The value loss for a shuffled mini-batch of level II is defined as follows:

$$\text{Value loss level II} = \sum_{k \in \mathcal{K}_2} L\left(V\left(s_k, \phi_2'\right) - u_k\right) \tag{3.23}$$

From the three different losses the total average loss of a mini-batch is calculated as follow:

$$\text{Average loss level I} = \frac{\text{Value loss level I} + \text{Policy loss level I} + \text{Entropy loss level I}}{B/\eta} \tag{3.24}$$

$$\text{Average loss level II} = \frac{\text{Value loss level II} + \text{Policy loss level II} + \text{Entropy loss level II}}{T \cdot B/\eta} \tag{3.25}$$

Minimising the average loss via updating θ' and ϕ' will result in the neural networks fitting to the given training data. The neural network parameters are iteratively updated using a stochastic gradient descent with the ADAM optimiser, as this optimiser is less sensitive to parameter tuning and is overall considered to be the current best practice (Kingma & Ba, 2014). The optimiser uses the average loss of minibatch j for actor and critic network of level h to calculate the gradient of the weights and bias with respect to the average loss and perform an update step with a learning rate of α. A low learning rate results in slow convergence to a good policy, while a high learning rate might results in overshooting a good policy.

Appendix 3.B Optimality gap

Table 3.7: Optimality gap of the PPO algorithm for different instances.

	Profit			Revenue	Costs			
Instance	MDP	PPO	Gap (%)	Gap (%)	Inventory Gap (%)	Ordering Gap (%)	Fulfillment Gap (%)	Return Gap (%)
$\mu_1 = 3, \mu_2 = 1$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$\mu_1 = 2, \mu_2 = 1$	1111.99	1088.10	−2.20	0.07	28.03	0.13	−0.20	−0.29
$\mu_1 = 1, \mu_2 = 1$	653.91	636.55	−2.73	0.37	28.07	0.30	0.53	0.62
$CV_i = 1/\sqrt{\mu_i}$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$CV_1 = 1/(2\sqrt{3})$	1616.12	1588.86	−1.72	0.02	24.80	0.01	0.04	0.05
$CV_2 = 1/(2\sqrt{1})$	1583.14	1554.56	−1.84	−0.07	22.68	0.03	−1.00	−1.00
$CV_i = 1/(2\sqrt{\mu_i})$	1626.19	1604.78	−1.33	−0.02	20.72	−0.02	0.06	0.06
$\rho_p = 0.2$	1676.88	1642.04	−2.12	−0.06	26.41	0.01	−1.22	−1.40
$\rho_p = 0.3$	1626.03	1589.67	−2.29	−0.19	22.69	−0.13	−0.83	−0.93
$\rho_p = 0.4$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$\rho_p = 0.5$	1526.03	1493.15	−2.20	0.03	26.53	0.05	−0.12	−0.12
$\rho_p = 0.6$	1476.48	1447.05	−2.03	0.25	23.45	0.12	0.96	0.98
$\rho_p = 0.2, \mu_2 = F(\rho_p)$	1590.26	1555.65	−2.22	0.07	26.79	0.14	−1.42	−1.52
$\rho_p = 0.3, \mu_2 = F(\rho_p)$	1587.48	1557.07	−1.95	−0.02	23.34	0.04	−0.78	−0.78
$\rho_p = 0.4, \mu_2 = F(\rho_p)$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$\rho_p = 0.5, \mu_2 = F(\rho_p)$	1558.63	1528.31	−1.98	0.17	23.23	0.08	0.62	0.68
$\rho_p = 0.6, \mu_2 = F(\rho_p)$	1540.07	1500.50	−2.64	0.97	32.12	0.66	2.24	2.21
$M = 1$	1576.78	1539.90	−2.40	0.08	26.83	0.07	0.32	0.18
$M = 2$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04

Continued on next page

Table 3.7: Optimality gap of the PPO algorithm for different instances (continued).

	Profit			Revenue	Costs			
Instance	MDP	PPO	Gap (%)	Gap (%)	Inventory Gap (%)	Ordering Gap (%)	Fulfillment Gap (%)	Return Gap (%)
$c_r = 0$	1588.93	1556.06	−2.11	0.10	24.74	0.08	0.28	0
$c_r = 5$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$c_r = 10$	1561.88	1524.68	−2.44	0.57	31.97	0.53	0.81	0.90
$c_p = 30$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$c_p = 50$	1078.26	1056.67	−2.04	0.88	24.97	0.82	1.47	1.44
$c_p = 70$	582.58	577.41	−0.89	3.67	22.76	3.06	9.30	9.41
$c_{h1} = 0.1, c_{h2} = 0.05$	1680.98	1667.44	−0.81	−0.02	55.71	0.05	−0.56	−0.61
$c_{h1} = 0.2, c_{h2} = 0.1$	1558.63	1528.31	−1.98	0.17	23.23	0.08	0.62	0.68
$c_{h1} = 2, c_{h2} = 1$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$L = 1$	1576.78	1539.90	−2.40	0.08	26.83	0.07	0.32	0.18
$L = 2$*	1575.15	1549.76	−1.64	0.17	21.60	0.18	−0.02	0.04
$L = 3$	1573.36	1537.44	−2.34	0.40	29.45	0.42	0.25	0.23
Average	1464.53	1432.82	−2.27	−0.44	23.09	−0.38	−1.09	−1.07

*Base test case with: $\mu_1 = 3, \mu_2 = 1, CV_i = 1/\sqrt{\mu_i}, \rho_p = 0.4, M = 2, c_r = 5, c_p = 30, c_{h1} = 1, c_{h2} = 0.5, L = 2$

Chapter 3

Appendix 3.C Goodness of fit

To evaluate how much the policy found by the PPO algorithm resembles the optimal policy, the weighted Normalised Root Mean Square Error (NRMSE) is used. The NRMSE is a common metric for comparison of DRL algorithms (e.g. Chi et al., 2010; Rocchetta et al., 2019; Xie et al., 2019). We simulate the heuristics for J periods where the set of periods is given by $\mathcal{J} = \{0, 1, \ldots, J\}$ and calculate the weighted NRMSE as follows:

$$\text{Weighted NRMSE} = \frac{\sqrt{\frac{1}{J} \sum_{j \in \mathcal{J}} \left(\pi_j^* - \pi_j\right)^2}}{\max_{j \in \mathcal{J}}(\pi_j^*) - \min_{j \in \mathcal{J}}(\pi_j^*)} \tag{3.26}$$

Here, π_j^* is the optimal action to be taken in simulation period j and π_j is the action chosen by the PPO algorithm policy found in the same period. Normalisation of the RMSE is performed by dividing the metric with the maximum and minimum value of the optimal policy, for cases where $\max_{j \in \mathcal{J}}(\pi_j^*) - \min_{j \in \mathcal{J}}(\pi_j^*) = 0$ the term is set to 1. For the ordering action the simulation period $J = 1.000.000$ periods and for the rationing action $J = T \cdot 1.000.000$ sub-periods.

Table 3.8: Weighted NRMSE of actions by PPO.

	Weighted NRMSE								
	Ordering	Rationing							
Instance		Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Average
$\mu_1 = 3, \mu_2 = 1$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.18
$\mu_1 = 2, \mu_2 = 1$	0.11	0.29	0.20	2.40	0.60	0.99	0.75	0.41	0.80
$\mu_1 = 1, \mu_2 = 1$	0.20	0.29	0.30	3.53	1.51	0.91	0.59	0.55	1.09
$CV_i = 1/\sqrt{\mu_i}$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.18
$CV_1 = 1/(2\sqrt{3})$	0.09	0.61	0.41	7.05	6.26	0.48	0.64	1.02	2.35
$CV_2 = 1/(2\sqrt{1})$	0.07	0.51	0.21	4.30	1.80	0.54	0.82	0.59	1.25
$CV_i = 1/(2\sqrt{\mu_i})$	0.11	0.51	0.33	3.69	2.33	2.90	0.33	0.46	1.51
$\rho_p = 0.2$	0.10	0.39	0.29	6.97	3.39	1.21	1.16	0.60	2.00
$\rho_p = 0.3$	0.09	0.23	0.19	4.02	2.79	0.61	0.65	0.41	1.27
$\rho_p = 0.4$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.18
$\rho_p = 0.5$	0.00	0.25	0.18	6.21	0.72	0.76	0.58	0.63	1.33
$\rho_p = 0.6$	0.00	0.35	0.26	3.92	0.72	0.69	0.45	0.40	0.97
$\rho_p = 0.2, \mu_2 = F(\rho_p)$	0.09	0.42	0.30	6.49	2.35	0.62	0.84	0.46	1.64
$\rho_p = 0.3, \mu_2 = F(\rho_p)$	0.15	0.23	0.13	4.57	3.07	0.72	0.78	0.44	1.42
$\rho_p = 0.4, \mu_2 = F(\rho_p)$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.18
$\rho_p = 0.5, \mu_2 = F(\rho_p)$	0.14	0.35	0.35	4.26	0.50	0.38	0.31	0.23	0.91
$\rho_p = 0.6, \mu_2 = F(\rho_p)$	0.14	0.17	0.24	9.88	0.94	0.72	0.50	0.16	1.80
$M = 1$	0.18	0.30	0.23	7.77	2.22	0.71	0.79	0.69	0.00
$M = 2$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.19

Continued on next page

Table 3.8: Weighted NRMSE of actions by PPO (continued).

	Weighted NRMSE								
	Ordering	Rationing							
Instance		Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Average
$c_r = 0$	0.00	0.21	0.21	5.93	3.47	0.68	0.94	0.49	1.70
$c_r = 5$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.19
$c_r = 10$	0.11	0.27	0.21	7.06	0.58	0.83	1.02	0.64	1.52
$c_p = 30$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.19
$c_p = 50$	0.10	0.18	0.21	3.28	0.48	0.69	0.50	0.43	0.82
$c_p = 70$	0.12	0.15	0.13	2.52	0.35	0.28	0.18	0.29	0.56
$c_{h1} = 0.1, c_{h2} = 0.05$	0.22	0.46	0.53	4.46	0.67	1.07	1.51	1.04	1.39
$c_{h1} = 0.2, c_{h2} = 0.1$	0.31	0.68	0.22	20.33	1.37	1.37	1.38	1.09	3.78
$c_{h1} = 2, c_{h2} = 1$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.19
$L = 1$	0.00	0.43	6.20	7.10	1.10	0.72	0.35	0.40	2.33
$L = 2$*	0.08	0.29	0.21	5.99	0.44	0.52	0.53	0.31	1.19
$L = 3$	0.00	0.30	0.23	7.77	2.22	0.71	0.79	0.69	1.82
Average of all instances	1.33	0.54	0.54	6.23	2.42	1.07	0.91	0.76	1.71

*Base test case with: $\mu_1 = 3, \mu_2 = 1, CV_i = 1/\sqrt{\mu_i}, \rho_p = 0.4, M = 2, c_r = 5, c_p = 30, c_{h1} = 1, c_{h2} = 0.5,$
$L = 2$

Chapter 4

Replenishment and Fulfilment Decisions for Stores in an Omni-channel Retail Network

This chapter is published as:

Goedhart, J., Haijema, R., Akkerman, R., & S. de Leeuw (*Under review*). Replenishment and fulfilment decisions for stores in an omni-channel retail network.

Abstract

Omni-channel retailing allows stores to be used to fulfil online orders. The allocation of online orders to stores can however be complicated, as one needs to take into account the inventory level of each store as well as potential future in-store demand. Current practice is therefore often to use myopic order allocation rules. However, such rules may cause inventory levels to become imbalanced across the retailers network, which might result in (expensive) excessive stocks for some stores while other stores face stock-outs. We therefore study the online fulfilment and replenishment decision of an omni-channel retailer with multiple stores to fulfil the online orders, considering future demand. In contrast to previous literature, we explicitly formulate it as a multi-period problem and formulate and solve it as a periodic Markov Decision Process (MDP). Each period (e.g., week) an ordering decision is made and replenishment happens after a lead time, while online orders are allocated at the end of each subperiod (e.g., day). As the problem easily becomes intractable with multiple stores, we find an approximation of the optimal policy by decomposing the MDP and applying a one-step policy improvement approach. In an extensive numerical study, we compare our policy with two well-known heuristics from the literature and practice. The results indicate that our approach outperforms the heuristics on both profit and service levels. Further analysis shows that our method is better at allocating the online orders to the stores, resulting in more balanced and less fluctuating inventory levels across the retailers network.

4.1 Introduction

Brick-and-mortar stores are experiencing increased competition from online retailers. Although online retailing can be a threat to these physical stores, it also enables new opportunities, as store retailers can also offer their products via online shopping channels. Retailers that integrate their physical store with their online shopping channel are referred to as omni-channel retailers (Verhoef et al., 2015). Furthermore, customers nowadays expect a uniform experience across channels (e.g. pricing, availability, and assortment), as customers often use all available channel during their shopping journey (Bijmolt et al., 2021). The advantage for the retailer in integrating the shopping channels is that the assets of one channel can be leveraged across the others.

In omni-channel retailing, stores are used for new roles, such as online fulfilment, pick-up of online orders, offering product information, and handling product returns (Hübner et al., 2022). As stores are then used to support these additional activities related to omni-channel retail operations, the retailer often needs to reconsider their store operations, for instance adjusting their inventory and assortment decisions. One of the main complexities is that retailers need to rethink their strategy: where previously stores were operated individually, now stores need to operate together as the online channel operates across stores. Although the complexity of store operations increases when integrating the online channel with the physical store, leveraging the assets of the store network for online sales provides new opportunities (Bayram & Cesaret, 2021).

In this paper, we focus on using physical stores for the fulfilment of online orders, which is one of the key aspects of omni-channel operations (Govindarajan et al., 2021). Using brick-and-mortar stores for the fulfilment of online demand is called a ship-from-store (SfS) strategy and has several benefits (Bendoly et al., 2007). For instance, an SfS strategy can reduce delivery time as stores are often located closer to customers compared to online fulfilment centers. A short delivery time from a local store will help to compete with large online retailers that operate multiple online fulfilment centers spread around a country to ensure same-day or next-day delivery, a brick-and-mortar retailer needs to invest in new multiple online fulfilment centers to offer the same service (McKinsey, 2021c). However, by leveraging their existing store network they can achieve the same delivery times without investing in new online fulfilment centers throughout the country. This is the motivation why some large US retailers increase the use of the SfS strategy (examples are given in Mahar et al. (2021)). Furthermore, as online demand is not region-dependent, different stores can be used to fulfil an online order. As a result, the retailer can be flexible in deciding which store to use to fulfil an online order, within the constraints of consumer delivery lead-time requirements. An additional advantage of using stores for online fulfilment is that demand of walk-in and online customers may be pooled at store level when store inventory is used for both types of demand (Goedhart et al., 2022).

Unfortunately, the online fulfilment decisions that arise when using stores for online sales are particularly challenging (Caro et al., 2020). Multiple trade-offs need to be made, of which most are complicated by the uncertainty of future sales. When deciding which store should fulfil an online order, the retailer has to take into account the costs of fulfilling the order, but also factors such as the inventory levels of the stores. As the store inventories are also used to serve in-store customers, there is a trade-off between fulfilling the online order from a specific store or reserving the product for a future walk-in customer in that store and use another store to serve the online customer. Current retail practices do not include all these factors and are often based on simple policies that only take into account current inventory levels or distance from the store to the consumer (Govindarajan et al., 2021). Making a wrong decision might cause inventory imbalances across the stores in the retail network. These inventory imbalances will negatively influence the retailer's profit, as some stores might end up with excessive stock, while others experience stock-outs.

Incorporating all information related to fulfilment decisions, such as inventory levels, outstanding orders, and future demands, can easily make the fulfilment problem intractable, making a well-informed decision difficult. Therefore, retailers often resort to simple rules by not taking into account potential future demand (Acimovic & Graves, 2015; Bayram & Cesaret, 2021). However, neglecting potential future demand may result in sub-optimal decisions that reduce profits and can lead to an inventory imbalance across stores, causing walk-in customers to encounter stock-outs. Thus finding a way to make a well-informed decision in which future demand is taken into account is a key element in successful retail operations.

Literature studying the online fulfilment decision often study settings where the retailer sells their product in a finite time period (e.g. Govindarajan et al., 2021). However, retailers also sell non-seasonal products that are sold throughout the year and replenished every period. Only limited studies have researched the online fulfilment decision for a retailer with multi-period replenishment (e.g., **?**), however their work does not optimise the replenishment decision or include a fixed lead time on the replenishment. While some previous work does take into account future demand in their online fulfilment decision (e.g., Govindarajan et al., 2021; Mahar et al., 2021), they do not take into account a multi-period replenishment decision with outstanding orders. To the best of our knowledge, there is no literature that tackles this important problem.

We study the ordering and online fulfilment decision of an omni-channel retailer operating multiple brick-and-mortar stores. We investigate how the stores can be utilised for the fulfilment of online orders. During the day, walk-in customers fulfil their demand with the products displayed on the shelves. At the end of the day, the retailer has to decide which of the different stores will fulfil the received online demand. Additionally, an ordering decision is made every week, which leads to an inventory replenishment after a fixed lead time.

We model the retailer's decision process as a Markov Decision Problem (MDP). Such problems easily become intractable as the number of states grows exponentially with the amount of stores. Therefore, we also propose an alternative, faster, method to solve the MDP in a near-optimal fashion, using an one-step policy improvement. This approximate solution is a method to get a good performing policy from a decomposed MDP. Decomposition of the MDP will be performed by solving the problem for individual stores separately and using their relative value vectors for the one-step policy improvement. The advantage of this method is that the information on future demand is still considered, and only the interaction between stores is approximated. We compare the performance of our proposed method with two commonly used heuristics.

We contribute to the scientific literature in several ways. First, we formulate a model for replenishment and allocation of online orders in an omni-channel retailers network context, in which stores are used for both physical and online demand. Second, we present a fast approximate method based on an exact solution approach to solve the allocation of online orders to stores. Contrary to previous literature (e.g., Acimovic & Graves, 2015; Govindarajan et al., 2021) our method incorporates current and future revenues and costs. Third, based on an extensive numerical analysis, we show that our method outperforms well-known existing heuristics on several performance indicators.

The remainder of this paper is structured as follows. Section 4.2 presents related research on online fulfilment via stores. In Section 4.3, we describe the decision problem in more detail and formulate it as an MDP. In Section 4.4, we introduce the one-step policy improvement method for the online order allocation and two well-known heuristics that have been discussed in previous literature and used in practice. In Section 4.5, we investigate the results of the one-step policy improvement solution and compare it with two heuristics. Section 4.6 concludes the paper and discusses future research directions.

4.2 Literature Review

The current literature about using stores for the fulfilment of online orders has been mostly focused on the strategic issue of defining which stores to use for fulfilling online orders (Ishfaq & Raja, 2018). One of the earliest studies investigating use of stores for online fulfilment is by Bendoly et al. (2007), in which the authors focused on whether using the stores or a dedicated fulfilment centre for online orders is ideal. They conclude that the fraction of online demand is an important factor for this strategic decision, but their study is limited by having the online orders evenly distributed across the stores.

Previous literature has paid less attention to the inventory and fulfilment decision, and especially the replenishment decision (Hübner et al., 2022). In the following, we discuss papers that consider a retail network consisting of stores with in-store demand, in which the retailer receives online demand in a central place and uses the network of stores to fulfil these online orders. We focus on papers that allocate online orders to stores, which is one of the key decisions that we study in this paper.

The first paper that included an online fulfilment decision is by Xiao et al. (2009). They use dynamic programming to derive an optimal policy for the allocation of online orders to two stores and compare it with a static policy where an online order is assigned to the closest store with inventory. The static policy results in imbalanced inventory levels between the two stores which results in significant revenue loss. The optimal policy keeps a better balanced inventory across the stores. However, as the state dimension easily becomes multidimensional in their study, calculating the optimal policy becomes too complicated for more than two stores. Mahar et al. (2009) and Bretthauer et al. (2010) similarly mention that for large problem instances, the computationally complexity makes the problem intractable. They mention that development of good performing heuristics is important. Acimovic & Graves (2015) mentions that current practice of most retailers is a myopic heuristic, in which potential future demand is not taken into account when making the online order fulfilment decision. Jia et al. (2021) investigates the online order fulfilment in a two-period setting, and found that at high inventory levels a myopic policy is effective but at low inventory levels a heuristic that takes into account future demand may help. They mention that such a heuristic, together with general multi-period models are interesting directions of future work. Bayram & Cesaret (2021) develop a finite-horizon MDP for the online fulfilment decision, and compare this with the common current retail practice mentioned by Acimovic & Graves (2015). Their method is limited as it is not suited for large-scale instances. Additionally, they mention that their assumption of fixed initial inventory levels is a limitation and future research should also include the inventory replenishment decisions.

One of the earliest studies that includes both an online fulfilment and replenishment decision is by Gupta et al. (2019). They model the ordering, fulfilment, and pricing decision for a retailer facing online and in-store demand as a multi-objective optimisation problem. They utilise a common search algorithm for finding the optimal policy, but also mention that future research could be focused on developing more suitable heuristics for the studied problem. Govindarajan et al. (2021) includes the initial replenishment decision, but to keep their problem tractable, they decouple it from the online fulfilment decision. They compare their method with the one described by Acimovic & Graves (2015) and a threshold-based myopic policy, in which inventory is reserved for future demand by not using the location if the inventory is below a time-dependent threshold. Although their policy outperforms the two myopic policies, their study does not consider multi-period replenishment, as they mention the complexity of multi-period models may result in them becoming intractable. The study by Mahar et al. (2009, 2021) does incorporate multi-period replenishment, however instead of optimising the (state-dependent) order quantities, they impose a base stock policy. Finally, the studies by Arslan et al. (2021) and Jiu (2022) do consider multiple ordering decisions, but also their study does not consider a general multi-period model.

Although the online order fulfilment problem has been extensively studied in the literature, no literature takes into account the replenishment problem. More specifically, they lack to optimise the replenishment decision for a retailer selling a product which is replenished via a regular replenishment cycle. Studies mention that including such a decision can make the model intractable due to the dimensionality of the problem, which increases exponentially with each store included in the retailers network. However, as stores normally have regular replenishment cycles (with a fixed lead time), the ordering decision is important. The replenishment decision and the online order allocation can influence each other, therefore optimising them simultaneously is important for better-aligned decisions. We contribute to the current literature by studying a retail store network, in which in-store demand is satisfied and consequently online orders are fulfilled from remaining store inventory. We explicitly include a multi-period replenishment decision where replenishment occurs after a fixed lead time. As indeed the state dimension easily becomes too large to be solved due to the number of stores in the network, we derive from an exact method an alternative method to find a well-performing policy that still incorporates both the ordering and fulfilment decision. We compare our method with two heuristics found in practice and literature.

4.3 Problem Formulation

We study the setting of an omni-channel retailer with a network of J physical stores, and describe the set of physical stores as $\mathcal{J}$ with $\mathcal{J} = \{1, 2, \dots, J\}$. The retailer has two different options for customers to fulfil their demand: offline and online. Here, offline means that the customer will visit the store to fulfil their demand, and online means that the customer orders online to have the product shipped to their home.

The retailer sets multiple decisions, which we model in subsequent time periods. Figure 4.1 depicts the different sub-periods within one period of the retailer. Every period the retailer determines the order quantity (q_j) for store j, which is replenished after a fixed lead time of L sub-periods. A period is divided into R sub-periods: at the beginning of the sub-period the demand d_j occurs and at the end the accumulated online orders (f_j) are allocated to the different stores.

We formulate the problem as a Markov Decision Process (MDP), in which the state of the system is dependent on the time of the sub-period. The state of the system $s = (I, Q, F, e, t)$ consists of the inventory level of each store, $I = (I_1, I_2, \dots, I_J)$, where

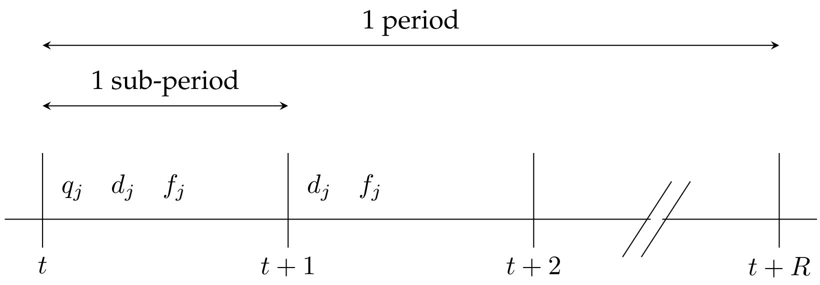


Figure 4.1: Visualisation of a period, sub-period and events.

I_j is the inventory level of store j in the omni-channel retail network. The outstanding orders of each store are denoted by $Q = (Q_1, Q_2, \ldots, Q_J)$, which we keep track of as these are replenished after a fixed lead time of L sub-periods. Furthermore, we keep track of the collected online orders that are received in the sub-period by the omni-channel retailer, which is denoted by F. Additionally, to be able to distinguish the start and end of the sub-period, the sub-period is split into two epochs, where $e = 1$ indicates it is the beginning of the sub-period and $e = 2$ the end. Lastly, we keep track of the sub-period denoted by t.

The objective of the MDP is to maximise profit, which is the result of the sales revenue and the different types of cost. These costs include holding and handling costs for the individual channels, online fulfilment costs, costs of not fulfilling an online order, and costs of the product. The costs differ among the different channels, however the price for the customer is equal in all channels, which is normal practice in omni-channel retailing (Gao et al., 2022).

The chronological order of events is presented in Figure 4.2. A square symbol indicates a state in which a decision is to be made, and a circle indicates a stochastic process. At the beginning of the sub-period, an ordering decision (q_j) is made, however only when it is also the beginning of the period. Then the demand of both channels realises (d_j). Here, in-store customers first fulfil their demand by purchasing products from store shelves, and the closing inventory of each store (I') is known. customers that encounter a stock-out in-store do not have their demand back ordered. At the end of the sub-period, the F online orders received are known. With the information about the closing inventory and online orders received, the online fulfilment decision per store f_j can be made. After the online orders have been fulfilled, the closing inventory is used for the next sub-period.

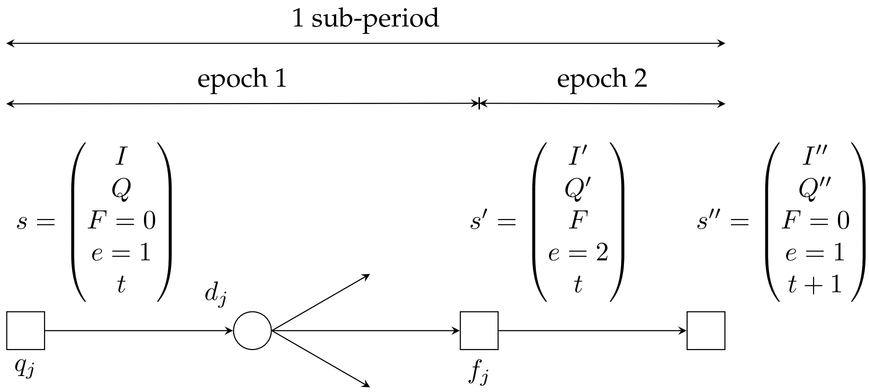


Figure 4.2: Visualisation of the chronological order of the states, actions, and stochastic process.

State transition during the sub-period: $s \rightarrow s'$:

The state at the start of the sub-period is $s = (I, Q, F = 0, e = 1, t)$. At the end of the sub-period ($e = 2$), the in-store demand is fulfilled, and the number of online orders F to be fulfilled is known. The new state of the MDP is: $s' = (I', Q', F, e = 2, t)$. I' is the inventory level of each store after demand of in-store has been fulfilled, Q' the outstanding order that does not change during the sub-period thus $Q'_j = Q_j$, F is the number of products accepted to sell via the online channel at sub-period t.

State transition to the next sub-period: $s' \rightarrow s''$:

After the collected online orders have been fulfilled from the stores, the closing inventory of the sub-period (I'') is used for the next sub-period. The state $s'' = (I'', Q'', F = 0, e = 1, t + 1)$ denotes the state of the system at the end of the sub-period, and is the new state s for the beginning of the next sub-period. As the outstanding order state does not change during the second epoch either, $Q''_j = Q'_j$. Additionally, all online orders that could not be fulfilled are considered to be lost thus $F = 0$ which is a common assumption (e.g., Acimovic & Graves, 2015; Bayram & Cesaret, 2021).

Demand

The demand in the different sales channels represents the uncertainty in the MDP. Similar to Govindarajan et al. (2021) and Bayram & Cesaret (2021), we assume offline demand is bound to a store and customers do not switch within the offline channel. The online demand is not store dependent and can be fulfilled from any store. A pooling effect is thus created as all combined store inventory can be used for online demand.

The demand in each channel is modelled according to a discrete demand distribution. The offline demand is given as d_j for all stores $j \in \mathcal{J}$. The probability of a demand for d_j items in store j is $P_j(d_j)$ with $d_j \in \mathcal{D}_j = \{0, 1, \ldots, D_j\}$. The demand of the online channel is given as $d_0 \in \mathcal{D}_0 = \{0, 1, \ldots, D_0\}$, with probability distribution $P_0(d_0)$. The maximum number of products to be distributed from store j on a given sub-period is thus $D_j + D_0$.

State space

The state space contains five dimensions: the inventory level of each store, the outstanding orders of each store, the collected online orders, the sub-period, and the epoch.

The state space of the inventory levels of each store in the network is limited by the maximum demand of the store during the review period plus lead time: $(R + L)(D_j + D_0)$. As a store could potentially fulfil all online demand, the maximum demand of a store is both the maximum in-store demand and maximum online demand. The inventory level of a store is thus: $I_j \in \mathcal{I}_j = \{0, 1, \ldots, (R + L)(D_j + D_0)\}$.

The state space of the outstanding orders depends on the sub-period t. After replenishment has occurred ($t > L$), the outstanding order is zero, and thus the state space is defined as $Q_j \in \mathcal{Q}_j(I_j) = \{0\}$. When $t \leq L$, the state space of the outstanding orders depends on the current inventory level and the total demand during review period and lead time and defined as: $Q_j \in \mathcal{Q}_j(I_j) = \{0, 1, \ldots, (R + L)(D_j + D_0) - I_j\}$.

The state space for the accepted online demand (F) is limited by the maximum online demand D_0: $F \in \mathcal{D}_0$.

The state space of the epoch consists of the beginning and end of the sub-period: $e \in \{1, 2\}$. The state space of the period is limited by the review period: $t \in \{1, 2, \ldots, R\}$.

Actions and action space

At the beginning of the period, a replenishment decision q_j is made for each store j, the combination of each store replenishment decision is given as $q = (q_1, q_2, \ldots, q_J)$. The action space of the replenishment decision is equal to that of the ordering state $q_j \in \mathcal{Q}_j(I_j)$. The ordering action space of each stores combined is given by $\mathcal{Q}(I) =$

$\prod_{j\in\mathcal{J}} \mathcal{Q}_j(I_j)$, thus the vector consisting of all order decisions has an actions space of $Q \in \mathcal{Q}(I)$.

At the end of each sub-period, the retailer has to allocate the collected online orders to the stores. As any store can fulfil any online order the allocation decision has to be made for each store, and we denote this decision as f_j. The combination of each store allocation decision is given as $f = \{f_1, f_2, \ldots, f_J\}$. The allocation decision is limited by the closing inventory of the store and the number of collected online orders, i.e. $f_j \in \mathcal{F}_j\left(I'_j, F\right)$ with $\mathcal{F}_j\left(I'_j, F\right) = \left\{0, 1, \ldots, \min\left(I'_j, F\right)\right\}$. Furthermore, f_j is restricted by $\sum_{j\in\mathcal{J}} f_j \leq F$, as no more online orders can be fulfilled than have been received. The allocation action space of all stores combined is given by $\mathcal{F}(I', F) = \prod_{j\in\mathcal{J}} \mathcal{F}_j\left(I'_j, F\right)$, thus the vector consisting of all allocation decisions has an actions space of $f \in \mathcal{F}(I', F)$.

State transitions

The state transitions at epoch one ($e = 1$) consists of the in-store demand being fulfilled, ordering, replenishment, and accepting online orders. The state transition for the inventory at each store for epoch one is dependent on the in-store demand and given as follows:

$$I'_j = (I_j - d_j)^+ \tag{4.1}$$

Here, we use the notation $x^+ = \max(0, x)$. The transition of outstanding orders is dependent on the sub-period. At the beginning of the period the outstanding order is set to the order quantity. For the remaining sub-periods, it is dependent on whether the replenishment is still outstanding, as we do not need to keep track of the outstanding order after lead time. The transition is as follow:

$$Q'_j = \begin{cases} q_j & \text{if } t = 1 \\ Q_j & \text{if } 1 < t \leq L \\ 0 & \text{Otherwise.} \end{cases} \tag{4.2}$$

Lastly, the accepted online demand of the sub-period is limited by the system-wide inventory level at the beginning of the sub-period and the quantity of online demand received. The reason it is limited by the system-wide inventory level at the beginning of the sub-period is because the retailer would not accept more online orders than they will be able to fulfil. The accepted online state is transitioned as follows:

$$F = \min\left(\sum_{j\in\mathcal{J}} I_j, d_0\right) \tag{4.3}$$

The state transitions at epoch two ($e = 2$) consists of the fulfilment of collected online orders and the inventory replenishment. The inventory state transition is as follows:

$$I_j'' = I_j' - f_j + \delta(t = L) \cdot Q_j' \tag{4.4}$$

Here, $\delta(x)$ denotes the Kronecker delta, which gives the value 1 if $x =$ True, otherwise 0. The outstanding order state does not change during the second epoch, thus $Q_j'' = Q_j'$. The collected online orders that are not fulfilled at the end of the sub-period are lost, therefore at the end of the sub-period the number of collected online orders $F = 0$. At the end of the sub-period, the state t is also increased by one, at the end of the period (when $t = R$) then it is set back to one.

Expected immediate reward

The goal of the MDP is to maximise the expected profit in the total network of stores. In the first epoch, the profit is the result of the revenue of products sold in-store and the costs of holding and ordering. The expected immediate reward of store j in epoch 1 given the current state, action, and demand is calculated as follows:

$$EC_j(I_j, q_j) = \begin{cases} p\left(\sum_{d_j<I_j} d_j \cdot P_j(d_j) + \sum_{d_j\geq I_j} I_j \cdot P_j(d_j)\right) & \text{(4.5a)} \\ -\, c_h \cdot I_j & \text{(4.5b)} \\ -\, c_p \cdot \delta(t=1) \cdot q_j & \text{(4.5c)} \end{cases}$$

The first term (4.5a) reflects the sales from all customers that purchased their product in-store, where p is the price of the product. The second term (4.5b) reflects the holding costs (c_h), which is determined at the beginning of each sub-period. The last term (4.5c) reflects the ordering costs, which are only incurred at the beginning of the period. The total expected immediate reward for state s can than easily be calculated by the sum over all individual stores $EC(s,q) = \sum_{j \in \mathcal{J}} EC_j(I_j, q_j)$

The total immediate reward in epoch 2 consists of profit from fulfilling online orders and a penalty cost for every online order that is not fulfilled:

$$C(s', f) = \begin{cases} (p - c_s) \sum\limits_{j \in \mathcal{J}} f_j & \text{(4.6a)} \\ - c_l \left(F - \sum\limits_{j \in \mathcal{J}} f_j \right) & \text{(4.6b)} \end{cases}$$

The first term (4.6a) reflects the revenue resulting from fulfilling online orders. As customers expect free shipping, the shipping cost (c_s) is incurred by the retailer and thus subtracted. The second term (4.6b) reflects the penalty cost (c_l) incurred for each online order that is not being fulfilled.

4.4 Solution Approaches

As the problem becomes larger with multiple stores, the problem cannot be solved within reasonable computation time. With each extra store in the omni-channel retail network, the inventory and outstanding ordering state grow in size. Additionally, the action space of both ordering and fulfilment also grow in size. This growth of problem size is referred to as the curse of dimensionality. To get an exact solution for the MDP with value iteration or policy iteration, one needs to enumerate multiple times over each state and action combination, which becomes infeasible for such large problems. For example, the base test case introduced below in Section 4.5 consists of approximately $1.0 \cdot 10^{66}$ different states. Even for a smaller case with only three stores in the network the number of states would still be $2.2 \cdot 10^{8}$ and infeasible to compute with value iteration. Additionally, for smaller cases, the online order allocation becomes less complex thus the difference between different policies becomes more negligible.

4.4.1 One-step policy improvement

To solve the problem in a reasonable time, we decompose the MDP followed by a one-step policy improvement step. This allows us to find an approximation of the optimal replenishment and online fulfilment policy for the stores in the omni-channel retail network. The advantage of the one-step policy improvement is that one only needs to perform a single iteration over each state and action combination instead of multiple iterations. This results in a large reduction in computational time, and previous research has shown that such an approximate solution will often result in a well-performing or even near-optimal strategy (Haijema & Hendrix, 2014). To find a well-performing approximate solution, the decomposition however needs to maintain the characteristics of the original problem. Bhatnagar & Lin (2019) used the one-step policy improvement for the joint transshipment problem, in which they decomposed the total network problem into independent single-location problems. We propose to use a similar approach, in which we decompose the MDP into problems for single stores, which can individually be solved via value iteration to obtain a relative state-value approximation. With the state-value approximation we can apply the one-step policy improvement to obtain a well-performing one-step improved policy (OSIP). Below we describe the procedure we use to find this OSIP.

We first solve the MDP with value iteration for each individual store to find the expected future profit over n consecutive sub-periods, thus for each store $j \in \mathcal{J}$, we can calculate the relative value matrix $v_{n,j}$. We use s_j to denote the state of an individual store j, thus $s_j = (I_j, Q_j, F, e = 1, t)$ and $s'_j = \left(I'_j, Q'_j, F, e = 2, t\right)$. As all stores share the online demand, we approximate the amount of online demand that each individual store receives by evenly splitting the average online demand (μ_0) over all stores in the retailers network and reshaping the probability distribution to fit the new average online demand given by $\widetilde{P}_0$.

The computation of $v_{n,j}$ is split into two stages corresponding with epoch 1 and 2. First we set $v_{0,j} = 0$. Next, one computes $v_{1,j}$, $v_{2,j}$, etc. using the recursive Bellman equations. First, the online order fulfilment decision is determined by computing $v_{n,j}\left(s'_j\right)$ for all states:

$$v_{n,j}\left(s'_j\right) = \max_{f_j \in \mathcal{F}_j(I'_j, F)} \left\{ C\left(s'_j, f_j\right) + v_{n-1,j}\left(s''_j\right) \right\} \tag{4.7}$$

To compute $v_{n,j}\left(s_j\right)$ for all states, the replenishment decision and in-store demand needs to be used:

$$v_{n,j}\left(s_j\right) = \max_{q_j \in \mathcal{Q}_j(I_j)} \left\{ EC_j(I_j, q_j) \sum_{d_j=0}^{D_j} \sum_{d_0=0}^{D_0} P_j\left(d_j\right) \widetilde{P}_0\left(d_0\right) \cdot v_{n,j}\left(s'_j\right) \right\} \tag{4.8}$$

Next, one continues computing $v_{2,j}$, $v_{3,j}$, etc. using the so-called recursive Bellman equations in equation (4.7) and equation (4.8). The sub-period t can be calculated from the iteration n as follow: $t = (R - (n-1) \mod R)$. The distinction of sub-period t is relevant to keep track of whether an order decision should be taken or an outstanding order has arrived.

We only apply the one-step policy improvement on the online fulfilment decision, for this we need the relative value matrix $v_{n,j}(s_j)$ of each store. To ensure that all $v_{n,j}(s_j)$ are computed with the same number of iterations n, n is set to a value n_{max} to ensure that the MDP of each individual store converges to a span below 0.1. Here, the span is defined as $\|v_{n,j} - v_{n-T,j}\|$. We can then transform the value matrix of the individual stores to approximate the expected future profit of the omni-channel retail network:

$$\widetilde{v}(s) = \sum_{j \in \mathcal{J}} v_{n_{max},j}\left(s_j\right) \tag{4.9}$$

With $\widetilde{v}(s)$ as the approximated expected profit when in state s. As we now have an approximation of the expected profit we can set the the average online demand back to its original value and use the original demand distribution $P_0(d_0)$. With the approximated expected profit, one can easily apply the one-step policy improvement to find the OSIP:

$$\widetilde{\pi}^f(s') = \operatorname*{arg\,max}_{f \in \mathcal{F}(s')} \left\{ C(s', f) + \widetilde{v_n}(s) \right\} \tag{4.10}$$

Here, $\widetilde{\pi}^f(s')$ is the approximated online fulfilment decision. Algorithm 4 gives an overview of the implementation of the one-step policy improvement. From the algorithm, the advantage of using the one-step policy improvement can be observed: there is no need to enumerate over all possible store state combinations for value iteration, instead value iteration can be applied consecutively for each store.

One could enumerate over all possible states to calculate the policy for each state with

Algorithm 4: Algorithm for the one-step policy improvement.

initialisation: $v_{0,j} = 0$; Set n_{max} so each value iteration converges to a span ≤ 0.1 ;
Set $\mu_0 = \mu_0/J$; Fit P_0 to μ_0 to form $\widetilde{P}_0$;
for $j \in \mathcal{J}$ **do**
  for $n = 1 : 1 : n_{max}$ **do**
    for $t = 7 : -1 : 1$ **do**
      for $s'_j = \left(I'_j, Q'_j, F, e = 2, t\right) \in \left\{\left(I'_j, Q'_j, F, e = 2, t\right) \mid I'_j \in \mathcal{I}_j, Q'_j \in \mathcal{Q}_j\left(I'_j\right), F \in \mathcal{D}\right\}$ **do**
        $v_{n,j}\left(s'_j\right) = \max_{f_j \in \mathcal{F}\left(I'_j, F\right)} \left\{ C\left(s'_j, f_j\right) + v_{n-1,j}\left(s''_j\right) \right\}$ where: s''_j is from equation (4.4)
      end
      for $s_j = (I_j, Q_j, F = 0, e = 1, t) \in \{(I_j, Q_j, F = 0, e = 1, t) \mid I_j \in \mathcal{I}_j, Q_j \in \mathcal{Q}_j(I_j)\}$ **do**
        $v_{n,j}(s_j) = \max_{q_j \in \mathcal{Q}_j(I_j)} \left\{ EC_j(I_j, q_j) + \sum_{d_j=0}^{D_j} \sum_{d_0=0}^{D_0} P_j(d_j) \widetilde{P}_0(d_0) \cdot v_{n-1,j}\left(s'_j\right) \right\}$ where:
        s'_j is from equation (4.1), (4.2) and (4.3)
      end
    end
  end
end
Set $\mu_0 = \mu_0 \cdot J$; **for** $t = 1 : 1 : 7$ **do**
  for $s = (I, Q, F = 0, e = 1, t) \in \{(I, Q, F = 0, e = 1, t) \mid j \in \mathcal{J}, I_j \in \mathcal{I}_j, Q_j \in \mathcal{Q}_j(I_j)\}$ **do**
    $\widetilde{v}(s') = \sum_{j \in \mathcal{J}} v_{n_{max},j}(s_j)$
  end
end
for $t = 1 : 1 : 7$ **do**
  for $s' = (I, Q, F, e = 2, t) \in \{(I, Q, F, e = 2, t) \mid j \in \mathcal{J}, I_j \in \mathcal{I}_j, Q_j \in \mathcal{Q}_j(I_j), F \in \mathcal{D}\}$ **do**
    $\widetilde{\pi}^f(s') = \arg\max_{f \in \mathcal{F}(I,F)} \left\{ C(s', f) + \widetilde{v}(s) \right\}$
  end
end

the one-step policy improvement; however, this procedure can still easily become too time-consuming as the number of states grows exponentially with each extra store. As we are only interested in the online fulfilment decision for relevant states, we simulate the omni-channel retailer network and only compute the online fulfilment decision via one step policy-improvement for the visited states. For the ordering decision, we use the ordering policy found by the value iteration of each individual store.

4.4.2 Heuristics

To evaluate the performance the OSIP, we will compare it with two heuristics found in the literature and retail practice. The first heuristic is a myopic policy in which the online fulfilment decision is made by only taking into account the expected contribution of the next sub-period without considering any more potential future in-store demand. The second heuristic determines the online fulfilment decision similarly, but also uses a threshold to protect some store inventory for future in-store demand.

Myopic fulfilment policy

Acimovic & Graves (2015) mentions that in practice, many retailers make their fulfilment decision only based on the current inventory level of each fulfilment location. Their decision is then based on minimising costs without taking into account any expected future costs. Based on this principle, we consider a myopic fulfilment policy (MFP).

In the MFP, we assume that when an online order needs to be fulfilled at the end of the sub-period, the retailer values the change in inventory level based on the highest opportunity cost of each store for the next sub-period. The opportunity cost is based on the expected demand of the next sub-period in-store customers. The opportunity cost of store j when fulfilling f_j online orders can be obtained using equation (4.5):

$$OC_j\left(I'_j, f_j\right) = EC_j\left(I'_j - f_j, 0\right) - EC_j\left(I'_j, 0\right) \tag{4.11}$$

Here, the term $EC_j(I'_j - f_j, 0)$ represents the expected contribution of store j with inventory level I'_j fulfilling f_j online orders, and $EC_j(I'_j, 0)$ represents the contribution when fulfilling no orders. From the opportunity cost of each store, the total opportunity cost when making the online fulfilment decision f can be calculated:

$$OC\left(s', f\right) = -c_l\left(F - \sum_{j\in\mathcal{J}} f_j\right) + (p - c_s)\sum_{j\in\mathcal{J}} f_j + \sum_{j\in\mathcal{J}} OC_j\left(I'_j, f_j\right) \tag{4.12}$$

The online fulfilment decision can be made by assigning online orders to the store with the highest opportunity cost. However, the retailer should not fulfil an online order when the opportunity cost is lower than the penalty cost, because then it is more

Algorithm 5: Algorithm for the myopic fulfilment policy.

for $t = 1 : 1 : 7$ **do**
  for $s' = (I', Q', F, e = 2, t) \in \{(I', Q', F, e = 2, t) \,|\, j \in \mathcal{J}, I'_j \in \mathcal{I}'_j, Q'_j \in \mathcal{Q}'_j(I_j), F \in \mathcal{D}\}$ **do**
    $\widetilde{\pi}^f(s') = \arg\max_{f \in \mathcal{F}(I,F)} \{OC(s', f)\}$;
  end
end

profitable to hold the product for potential in-store customer in the next sub-period and incur the cost of cancelling an online order. Algorithm 5 gives an overview how the heuristic policy is derived.

The advantage of Algorithm 5 compared to Algorithm 4 is that it skips the computationally expensive value iteration for finding the optimal online fulfilment decision. Furthermore, one can use the monotonicity of the outstanding orders to its advantage as the policy of state $\widetilde{\pi}^f\left(s' = \left(I', Q'_1 = 0, \ldots, Q'_j = 0, F, e = 2, t\right)\right)$ is equal to $\widetilde{\pi}^f\left(s' = \left(I', Q'_1 = 1, \ldots, Q'_j = 0, F, e = 2, t\right)\right)$ and $\widetilde{\pi}^f\left(s' = (I', Q'_1 = 2, \ldots, Q'_j = 0, F, e = 2, t\right)\right)$, etc.

Threshold-based myopic fulfilment policy

The myopic policy described above tries to fulfil all online demand of the sub-period. Govindarajan et al. (2021) discusses that doing so may be suboptimal as you might lose more profitable future in-store demand. Therefore, they mention that rationing the store inventory to hold back part of the stock for future in-store customers might be more profitable. Based on this principle, we extend the MFP to include a rationing decision that determines a threshold for the individual store inventory. This threshold inventory is then held back for future in-store customers. We call this a threshold-based myopic fulfilment policy (TMFP). Unlike the study of Govindarajan et al. (2021), we do assume a replenishment cycle, and we therefore adapt their rationing strategy so that it does not consider demand up until the end of a selling season, but until replenishment occurs. The number of products that are held back each sub-period for future in-store customers is calculated by balancing the future holding cost with the expected revenue from in-store demand:

$$w_j(t) = P_{j,t}^{-1}\left(\frac{p}{c_h\left((t-L) \mod R\right) + p}\right) \tag{4.13}$$

Here, $w_j(t)$ is the number of products the retailers holds back in store j in sub-period t. $P_{j,t}^{-1}$ is the inverse of the cumulative distribution of in-store demand from period t up until replenishment. The term $c_h\left((t-L) \mod R\right)$ reflects the maximum holding cost of a product that is reserved for future in-store customers.

To include the threshold in the fulfilment decision, the action space of the allocation of online orders is adjusted as follows:

$$\mathcal{F}_j\left(I'_j, F, t\right) = \left\{0, 1, \ldots, \min\left(\left(I'_j - w_j\left(t\right)\right)^+, F\right)\right\} \tag{4.14}$$

Note that the action space is now dependent on the sub-period, whereas with our approach or the MFP policy it was not.

4.5 Results

We evaluate the performance of the OSIP by evaluating it against the two heuristics for a wide range of instances. We first introduce a base test case, and compare the policies on economic performance, alpha service levels, and inventory performance. We then investigate the policies on a wider range of instances, focusing on profit, cycle service level, and the inventory balance before replenishment.

4.5.1 Data and design of experiments

The data set used is based on various recent literature studying similar omni-channel retailer network settings (Xu & Cao, 2019; Bayram & Cesaret, 2021; Goedhart et al., 2022). We assume that stores have varying in-store demand but all have identical review periods, lead time, price, and costs, which is often the case in practice (e.g., Noordhoek et al., 2018). We assume days as sub-periods and a review period of 7 days ($R = 7$) to reflect a weekly ordering decision, with a fixed lead time of $L = 2$, indicating that orders are placed on Monday ($t = 1$), and delivered on Wednesday ($t = 3$). We set the price of the product at $p = 100$. The different instances are created by varying subsets of parameters, which can be found in Table 4.1.

The stores differ on the average in-store demand: for the base test case we assume there are small stores with demand $\mu_j = 2$, medium stores with demand $\mu_j = 4$, and large stores with demand $\mu_j = 6$. Furthermore, for the base test case, the average

Table 4.1: Parameter values, where the base case is given in bold.

Parameter	Value	Values
Online demand	μ_0	$30, \mathbf{60}, 120$
In-store demand	μ_j	$\mathbf{2}, \mathbf{4}, \mathbf{6}$
Number of stores	m_k	$1, 2, 5, 7, \mathbf{10}, 11, 13, 15$
Procurement cost	c_p	$30, 50, \mathbf{70}$
Shipping cost	c_s	$5, \mathbf{10}, 20$
Lost sale cost	c_l	$0, 5, \mathbf{10}$
Holding cost	c_h	$0.25, 0.5, \mathbf{1}$

online demand is $\mu_0 = 60$. The offline and online demand is assumed to be independently Poisson distributed, and as we need a finite support they are right-truncated at a cumulative probability of 99.9% as described in Cohen (1954).

For the base test case, we assume there are 3 different types of stores, and we define each type of store as $k \in \mathcal{K}$ with $\mathcal{K} = \{1, 2, 3\}$. The number of each type of store is denoted by m_k and thus the total number of stores is $J = \sum_{k \in \mathcal{K}} m_k$. The number of small stores is $m_1 = 10$, the number of medium stores is $m_2 = 10$, and the number of large stores is $m_3 = 10$. We denote the number of stores with $M = \{m_1, m_2, m_3\}$.

The cost of the product is $c_p = 70$, the shipping cost of an online order is $c_s = 10$, the penalty of a lost sale is $c_l = 10$, and the holding cost per day is $c_h = 1$. In the remainder of this paper, if an instance does not specify a certain parameter value, it will be equal to their base test case value. The results were obtained by implementing the one-step policy improvement, and two heuristics in Python version 3.7.2. The model was run on a Personal Computer with Intel Xeon W-2133 CPU @ 3.60 GHz and 32GB of RAM. The economic performance, alpha service levels, and inventory levels were obtained by simulating the policies of all algorithms for 100,000 periods with a small warm-up period. As we are only interested in the difference between the online order fulfilment policies of the different approaches, the order policy obtained from the one-step policy improvement is also used for the two myopic heuristics.

4.5.2 Base test case

Economic performance

First, we evaluate the base test case on all the economic indicators. We split the weekly average profits, revenues, and costs for the three different types of store. We define the

Table 4.2: Average economic performance of the small store, medium store, large store, and the total economic performance of the network.

			Profit	Revenue	Cost			
					Holding	Fulfilment	Ordering	Penalty
Store 1		OSIP	552.24	2703.21	123.14	135.51	1892.25	-
	Gap (%)	MFP	0.82	-0.32	-7.08	1.15	-0.32	-
		TMFP	0.93	-0.22	-7.15	1.32	-0.22	-
Store 2		OSIP	912.51	4116.25	181.84	140.46	2881.38	-
	Gap (%)	MFP	-0.87	-1.06	-2.89	0.10	-1.06	-
		TMFP	-0.77	-0.99	-2.89	0.03	-0.99	-
Store 3		OSIP	1273.25	5520.14	240.55	142.18	3864.10	-
	Gap (%)	MFP	-1.43	-1.40	-1.33	-1.26	-1.40	-
		TMFP	-1.34	-1.34	-1.34	-1.38	-1.34	-
Total		OSIP	27379.96	123396.04	5455.27	4181.45	86377.19	2.16
	Gap (%)	MFP	-0.79	-1.05	-3.14	-0.02	-1.05	40.50
		TMFP	-0.69	-0.98	-3.17	-0.03	-0.98	62.11

gap as the percentage deviation from the OSIP. A positive percentage thus indicates that the heuristic has a higher value for the given performance indicator. Table 4.2 presents the economic performance for the OSIP, MFP, and TMFP.

From Table 4.2 it is observed that OSIP has higher total profits compared to the MFP and TMFP. However, the profit for small stores is lower under OSIP compared to the MFP and TMFP. Furthermore, it is observed that the TMFP outperforms the MFP in profit. Although the MFP and TMFP have higher profit for small stores, the gap for large stores is more significant. The result is that OSIP overall outperforms the MFP and TMFP.

The revenue for OSIP is higher across all stores compared to the MFP and TMFP, indicating that more products are being sold. As the order policy is equal for all policies, the number of products available in the network should be equal. However, if the inventory is unbalanced across the network, some in-store demand might not be satisfied due to stock-outs resulting in excessive stock at some stores and lost sales at others which can negatively influence the revenue.

The average holding cost is lower for the MFP and TMFP, this can be the result of lower order quantities, thus lower inventory levels. As the ordering costs are lower, this indicates that one has lower inventory holding cost. Furthermore, as small stores are more used for online fulfilment under the MFP and TMFP, they have less products in-store therefore lower holding cost.

The total fulfilment costs of the different algorithms are almost equal, indicating that the same amount of online demand is fulfilled. There is a difference in fulfilment cost among the different stores under the MFP and TMFP, where there is a preference to use the inventory of small stores to fulfil online demand as the cost is higher. Especially TMFP, where store inventory is protected for future in-store customers prefers to protects its inventory of large stores for these type of customers. OSIP show less preference for a store for the fulfilment of online orders, as the cost is similar across the three different type of stores.

Although the order policy of all three algorithms is equal, the MFP and TMFP order fewer products than OSIP. A numerical investigation found that at low inventory levels, the optimal ordering policy follows a base-stock policy, and at higher inventory levels it follows an order up-to policy. This suggests that if the inventory level of the store is below a certain level, the order quantity should be equal for all different inventories as an equal number of products is ordered under a base-stock policy. However, as the ordering costs are lower under the MFP and TMFP, this suggests that the inventory distribution at the beginning of the week is skewed to higher inventory levels. With higher inventory levels an order up-to policy is followed leading to lower order quantities.

The penalty cost is not split over the different types of store, as the results of an unfulfilled online sale cannot be assigned to a specific store. The results show that only 0.1% of online orders are not fulfilled under the OSIP, which results in a cancellation of an online order only once or twice per month. The MFP and TMFP have higher penalty costs (20-30 times that of the OSIP), but the penalty costs are still relatively low compared to all other costs. The TMFP has higher penalty costs than MFP as its threshold principle is basically there to protect in-store customers to some extent from online customers, thus rejecting online sales earlier.

Alpha service level

Besides profits, an important performance indicator is the alpha service level, which is the fraction of time one ends a day with products still in stock (i.e., the non-stock out probability). We report these for each store and the online channel for the three different online fulfilment policies in Figure 4.3.

The cycle service level, which is the alpha service level just before a replenishment

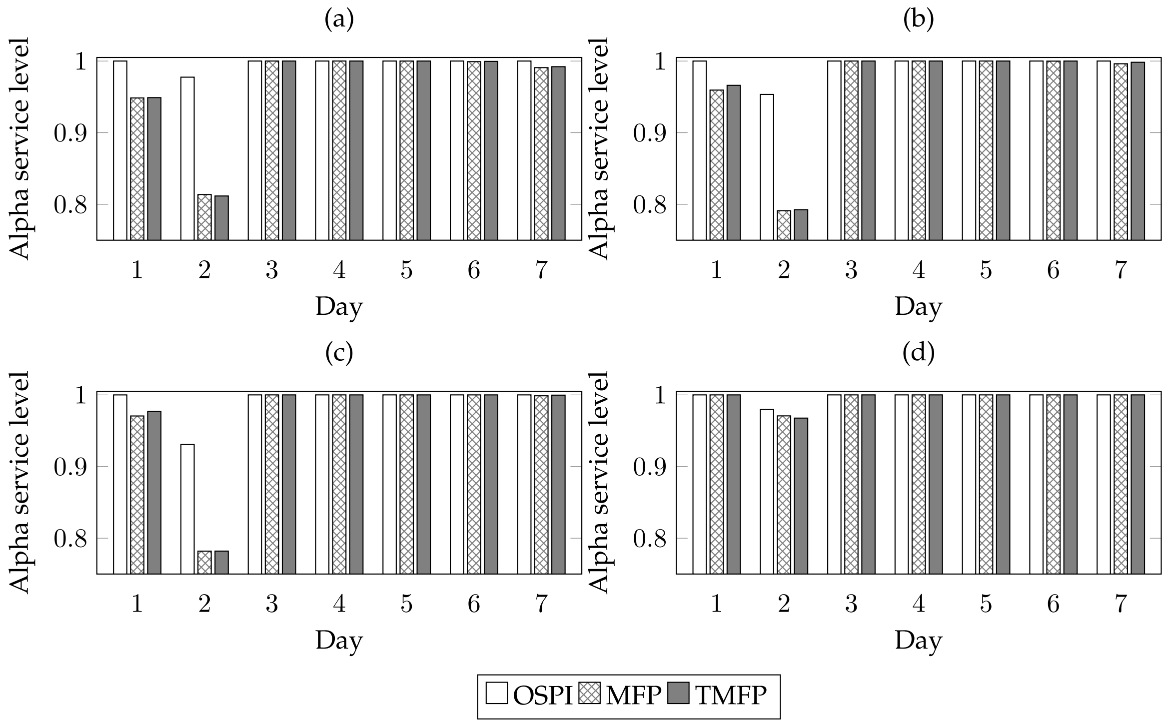


Figure 4.3: Alpha service level of small store (a), medium store (b), large store (c), and the online channel (d).

arrives, is the most interesting. It is observed that OSIP has a higher cycle service level for all stores and the online channel. Although Table 4.2 showed only slightly higher profits for OSIP, the alpha service level is significantly higher than it is for the MFP and TMFP.

Figure 4.3(a) shows the alpha service level of a small store, where it is observed that on day seven the alpha service level for the MFP and TMFP is already below 1. This is expected as it was observed that small stores are more used for fulfilment of online orders under these policies, thus their inventory depletes faster. The cycle service level of the online channel is almost equal to 1, which is expected as a stock-out across the whole retail network has a low probability. Figures 4.3(b) and 4.3(c) show the alpha service level of a medium and large store respectively. They show a similar trend as the small store, where on day seven the alpha service level is already just below 1 for the MFP and TMFP. For medium and large stores, TMFP has a higher alpha service level than MFP, which is expected as the TMFP protects their inventory for future in-store customers. This results in TMFP having a lower alpha service level for the online channel on the day before replenishment.

Although the order policy used in all three approaches is equal, the alpha service level for the MFP and TMFP is lower for some of the days and types of stores. As also seen in the results relating to the order cost, although a similar order policy is used, the online order fulfilment decision clearly influences the inventory levels through the selling period, resulting in imbalanced inventory levels across the retail network.

Inventory level

As observed from the economic performance and alpha service level, the online order fulfilment decision influences the inventory level of the store. As an unbalanced inventory across stores can lead to lost sales while others might have inventory, the online order fulfilment decision should be taken carefully. To better visualise how the inventory level differs between the different policies, we visualise the spread of the inventory level. Figure 4.4 shows the box plots of the inventory at the beginning of the day for each method and for each type of store. The average inventory level is given by an $\times$ mark. The box indicates the first and third quartile of the inventory level, with the median indicated by a line. The whiskers cover 95% of the data set.

From Figure 4.4, it is observed that the inventory level under the OSIP is more stable, as the whiskers and quartiles are closer to each other. A more fluctuating inventory level indicates more imbalance between stores, which can lead to stock-outs for some stores while there are still products available in the network. This also explains some of the results seen for the alpha service level, which was lower under the MFP and TMFP even though they use the same order policy as for the OSIP.

Furthermore, the lower ordering cost for the MFP and TMFP can be explained via the box plots. If the inventory level at the moment of ordering is below a certain level, a base stock policy is used. A numerical investigation showed that the threshold of the TMFP is around the median inventory level for all stores, thus inventory levels below the median order equal quantities. However, we notice in Figure 4.4 that the MFP and TMFP also have longer tails to higher inventory levels, for which the ordering policy follows more of an order up-to policy and therefore ordering less. Therefore, the longer tails to higher inventory levels result in lower order quantities but the tails to lower inventory levels do not result in higher order quantities, overall resulting in a total lower ordering cost.

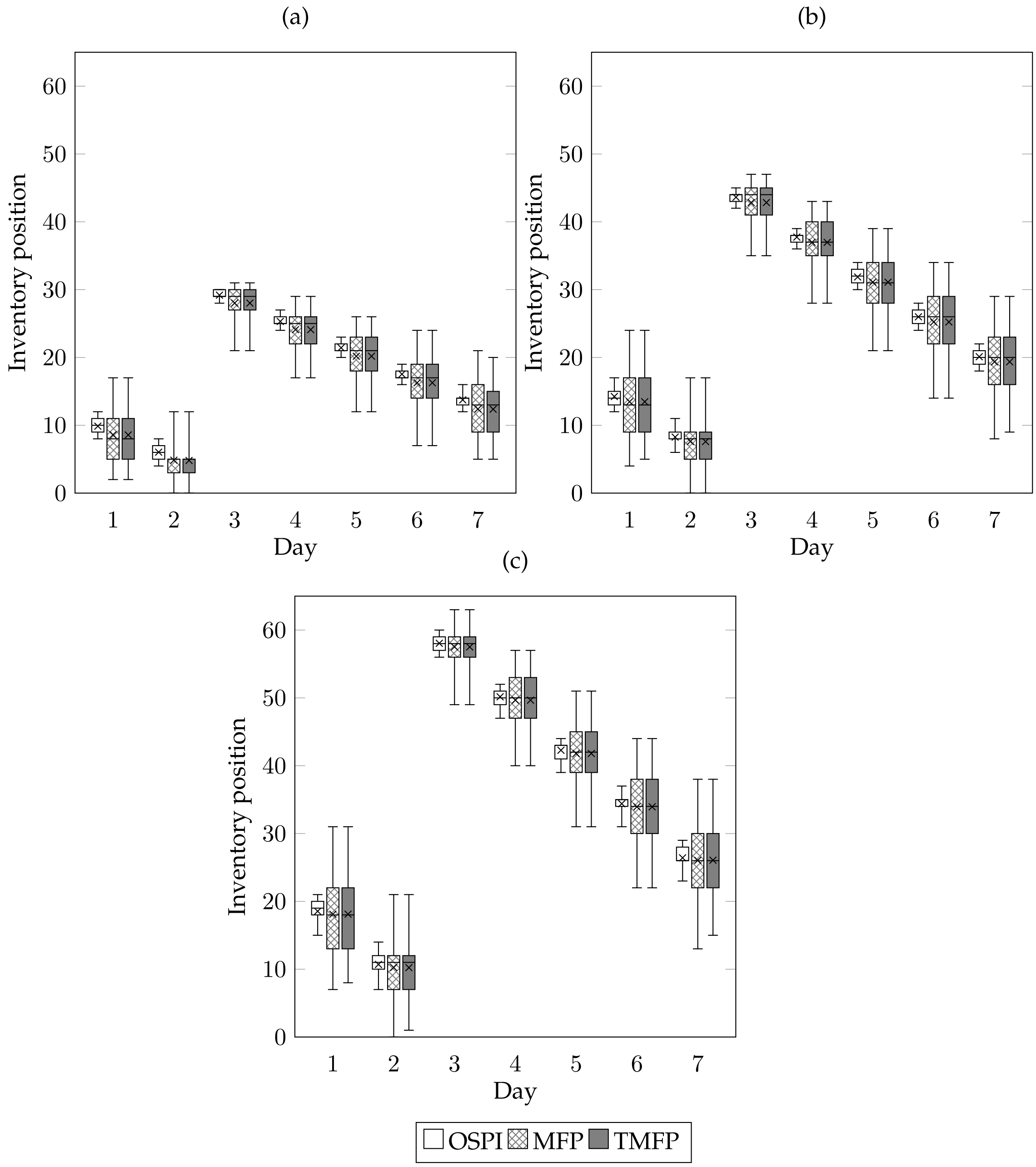


Figure 4.4: Inventory level of small store (a), medium store (b), and large store (c).

Fulfilment decision per day

To get better insight in how the inventory level of the OSIP is more stable (or less fluctuating) compared to the MFP and TMFP, we further compare the online order fulfilment decision per day of the policies. Figure 4.5 displays the average fulfilment costs per day for each type of store under each policy. The fulfilment costs indicates how many online orders are fulfilled from the different types of stores.

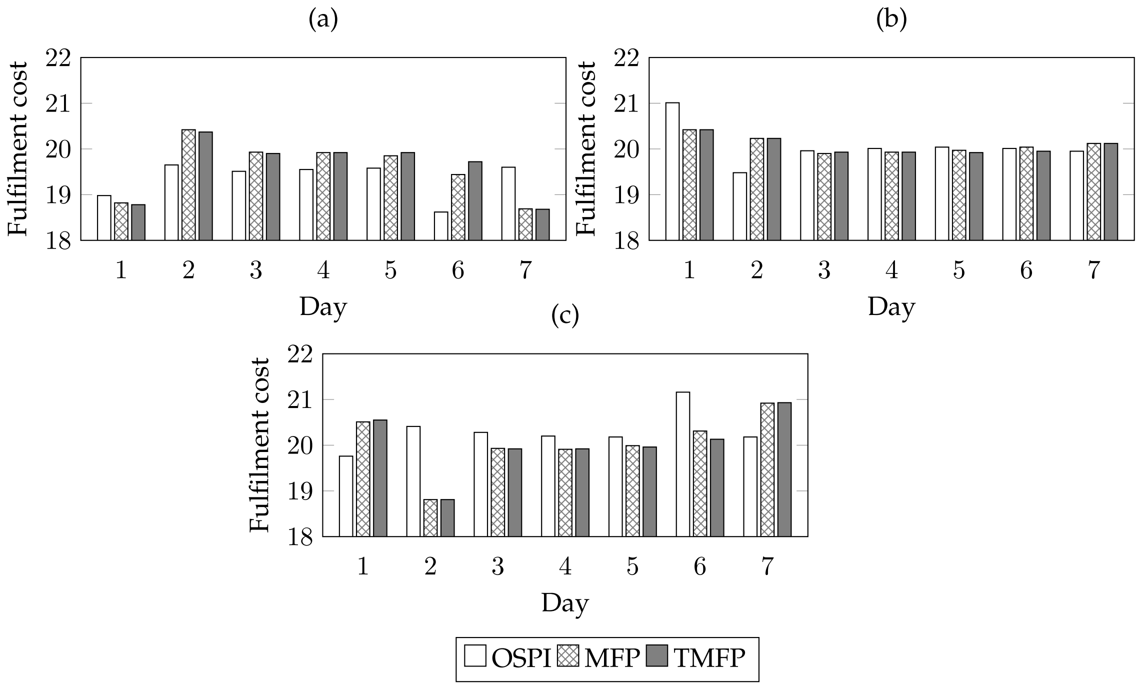


Figure 4.5: Fulfilment cost of each day of the small store (a), medium store (b), and large store (c).

Chapter 4

From Figure 4.5, it is observed that for the days after replenishment (day 3 till 5), the MFP and TMFP have almost equal fulfilment cost per type of store, indicating that all stores fulfil the same number of online orders. As the inventory levels during these days are high enough to fulfil the demand of the next day, the myopic online fulfilment decision of the MFP and TMFP remains equal as there is no trade-off between fulfilling an online order or fulfilling next-day demand. The OSIP has a preference for the larger stores when inventory levels are high.

On day 6 under the OSIP, large stores fulfil more online orders, likely to ensure that small stores has enough products available for the following days. The MFP and TMFP follow a similar pattern but to a lesser degree. As they protect small stores less, it is observed that on day 7 and 1, the MFP and TMFP have low inventory levels for small stores, thus use them less to fulfil online orders. This results in mostly large stores being used to fulfil online orders on these days, resulting in a low inventory level on day 2. Therefore on the day before replenishment small stores fulfil most of the online orders, as the other stores have less inventory available.

For OSIP on day 2, when the inventory is low for all stores, we observe that the OSIP uses large stores mostly to fulfil online orders. The reason for this can be two-fold.

First, large stores probably have large outstanding orders due to larger demand, thus will have enough products in stock for the next day. Second, as seen in Figure 4.4, large stores have a higher inventory level at day 2, and thus are therefore capable of fulfilling more online orders.

We can conclude from Figure 4.5 that due to the fact that the MFP and TMFP do not look ahead, inventory becomes unbalanced across the network. Although the TMFP does incorporate reservation for future demand, we observe it is not capable of making a good trade-off between fulfilling online orders or future demand as it performs similar to MFP. The OSIP is capable of making better informed fulfilment decisions, and therefore the policy is not equal across all days. We notice that on specific days, the inventory in specific types of stores are preferred to reserve the inventory of the other stores for future demand.

The results of the base test case demonstrate that OSIP outperforms the MFP and TMFP on economic performance, service level, and inventory uncertainty. Although the profit is only slightly higher in comparison with the MFP and TMFP, the cycle service level is much higher for OSIP. Reasons could be found in the fluctuation of the inventory level, which is much lower under the OSIP. This can be explained by the solution technique of one-step policy improvement, which better approximates the value of the next state by looking further than only the direct expected contribution. Therefore, by weighing the direct and future contribution better one will less likely end up in an unfavourable state in which the inventory level of the stores in the network becomes unbalanced.

4.5.3 Full set of instances

The results of the OSIP, MFP, and TMFP are evaluated for a wide range of instances as described in Section 4.5.1. We evaluate them on their economic performance, alpha service level, and variation in their inventory level. The first performance measure gives us an indication how well the OSIP performs on profit for each type of store, the second performance measure gives an indication on the performance of the demand fulfilment, and the last performance measure tells us something about how good the retailer can balance their inventory level.

Economic evaluation

Table 4.3 presents the profit for each type of store for all instances. It is observed that OSIP outperforms the MFP and TMFP for all instances. The MFP has a gap of around 0.70%, while the TMFP maintains a gap of around 0.60%. For almost all instances, small stores have a higher profit under the MFP and TMFP. Only for instances with a higher shipping cost or lower penalty for lost sale, OSIP has a higher profit for small stores.

When varying the number of each type of store, while maintaining an equal number for each type, the OSIP starts to perform better than the MFP and TMFP. However, when increasing the number of stores, the profit gap of small stores also starts to increase while for medium and large stores it decreases. Although the total profit increases with more stores due to more demand, the profit per store decreases as the total amount of online demand remains equal thus the fraction of online demand a store fulfils decreases. With fewer stores, taking into account future inventory levels of stores becomes less important for the fulfilment decision as the gap between OSIP and the MFP and TMFP becomes closer.

The instances with varying and unequal numbers of store per type reflect instances in which the average in-store demand remains equal while the number of each type of store is made unequal. The total profit slightly increases with more large stores, however the performance of the OSIP compared to the MFP and TMFP decreases. This is opposite from what we saw for the instances in which the number of stores per type was equal, as here the number of total stores increases, which has a negative effect on the performance of OSIP. The profit per store for all types also increases when having more large stores, which can be explained by the increase in share of online demand that each store has to fulfil. With an increase in large stores, the total number of stores decreases. The MFP and TMFP have a higher profit for small stores compared to OSIP, where the gap increases with a decreasing share of small stores in the network. A similar trend can be observed for the medium and large stores.

With a lower online demand, the inventory level of all stores is lower thus the online order fulfilment decision becomes more important as stock-outs can occur more frequently. In this instance, the gap between OSIP and the MFP and TMFP increases, indicating that OSIP makes better online fulfilment decisions. With higher online demand, the probability of stock-out decreases as stores will order more to fulfil the po-

Table 4.3: Profit of all instances.

	Small store			Medium store			Large store			Total		
		Gap (%)			Gap (%)			Gap (%)			Gap (%)	
	OSIP	MFP	TMFP	OSIP	MFP	TMFP	OSIP	MFP	TMFP	OSIP	MFP	TMFP
Base test case*	552	0.82	0.93	912	-0.87	-0.77	1273	-1.43	-1.34	27379	-0.79	-0.69
$M = \{1, 1, 1\}$	2506	0.10	0.10	2863	-0.43	-0.43	3224	-0.80	-0.79	8593	-0.41	-0.41
$M = \{2, 2, 2\}$	1425	0.03	0.04	1783	-0.50	-0.49	2144	-0.86	-0.87	10707	-0.50	-0.50
$M = \{5, 5, 5\}$	774	0.64	0.65	1132	-0.55	-0.53	1495	-0.89	-0.88	17017	-0.43	-0.42
$M = \{15, 15, 15\}$	471	1.67	1.86	831	-1.19	-1.04	1199	-2.20	-2.07	37533	-1.14	-0.99
$M = \{19, 10, 7\}$	518	0.19	0.34	877	-1.18	-1.05	1244	-2.16	-2.05	27330	-1.00	-0.87
$M = \{13, 10, 9\}$	534	1.41	1.53	901	-1.39	-1.29	1259	-1.49	-1.39	27296	-0.72	-0.62
$M = \{7, 10, 11\}$	567	1.08	1.21	928	-0.77	-0.67	1290	-1.30	-1.21	27442	-0.78	-0.68
$M = \{1, 10, 13\}$	607	1.52	1.66	960	0.02	0.11	1325	-0.93	-0.84	27445	-0.54	-0.45
$\mu_0 = 30$	431	1.09	1.37	795	-1.68	-1.51	1159	-2.24	-2.10	23868	-1.45	-1.28
$\mu_0 = 120$	778	0.71	0.73	1137	-0.37	-0.36	1500	-0.78	-0.76	34162	-0.30	-0.29
$c_h = 0.25$	635	1.98	1.99	1030	0.43	0.43	1448	-1.24	-1.24	31139	-0.03	-0.03
$c_h = 0.5$	603	1.93	1.96	990	-0.33	-0.30	1382	-1.18	-1.17	29759	-0.27	-0.25
$c_p = 30$	1600	2.63	2.64	2504	0.42	0.43	3473	-1.76	-1.76	75778	-0.11	-0.11
$c_p = 50$	1071	2.41	2.44	1710	-0.24	-0.22	2354	-1.23	-1.20	51358	-0.14	-0.12
$c_s = 5$	620	1.29	1.42	971	-0.28	-0.19	1340	-1.43	-1.36	29323	-0.47	-0.38
$c_s = 20$	421	-0.86	-0.81	778	-2.07	-1.95	1139	-2.42	-2.29	23393	-2.02	-1.91
$c_l = 0$	554	0.81	0.93	913	-1.29	-1.18	1277	-1.88	-1.78	27456	-1.14	-1.03
$c_l = 5$	552	0.91	1.04	912	-0.95	-0.84	1273	-1.63	-1.54	27382	-0.89	-0.79
Average	801	1.07	1.16	1207	-0.70	-0.62	1621	-1.47	-1.40	29704	-0.69	-0.62

*$M = \{10, 10, 10\}$, $\mu_0 = 60$, $c_h = 1$, $c_p = 70$, $c_s = 10$, $c_l = 10$

tential online orders. Therefore the online fulfilment decision becomes less important and the gap between OSIP and the MFP and TMFP becomes close to zero. The profit increases with the same number for each type of store, this indicates that the extra online orders are equally distributed over each type of store.

With a lower holding cost, the performance gap between OSIP and the MFP and TMFP decreases. With a lower holding cost, the number of products the retailer orders increases as the cost of having excessive stock decreases. Therefore, there are more products in the retailer's network and thus the online fulfilment decision becomes less important as the probability of a stock-out decreases. A similar trend can be observed for lowering the procurement cost: as this decreases, the retailer will order more products. In both these instances, small and medium store have a higher profit under the MFP and TMFP compared to OSIP. The MFP and TMFP prefer to use small and medium stores to fulfil online orders compared to OSIP resulting in a higher profit due to higher revenue and lower holding cost however increasing the risk of a stock-out at the end of the period.

As expected, a higher shipping cost decreases the profit for all types of stores. With a higher shipping cost, the gap between the OSIP and the MFP and TMFP increases. With a higher shipping cost, the profit margin per product decreases resulting in the retailer ordering fewer products to fulfil demand. Therefore, with lower inventory levels, the online fulfilment decision becomes more important, and thus the better performing OSIP has a higher profit compared to the MFP and TMFP. A lower penalty cost has little effect on the total profit, as the penalty cost is relatively small compared to other costs as seen in Table 4.2.

Cycle service level

In Table 4.4 we present the cycle service level for all instances. It is observed that for almost all cases the cycle service level of the policy of OSIP is higher than the MFP and TMFP. Only for the instance with an increased shipping cost, the online channel has a lower cycle service level with OSIP. The service level of each store type is almost equal under the MFP and TMFP for all instance.

From Table 4.4 it can be observed that with an increase in the number of stores, the cycle service level of each store becomes lower, while the cycle service level of the online channel increases. With more stores in the retailer's network, the expected number of

Table 4.4: Cycle service level of all instances.

	Small store			Medium store			Large store			Online		
	OSIP	MFP	TMFP	OSIP	MFP	TMFP	OSIP	MFP	TMFP	OSIP	MFP	TMFP
Base test case*	0.98	0.81	0.81	0.95	0.79	0.79	0.93	0.78	0.78	0.98	0.97	0.97
$M = \{1, 1, 1\}$	1.00	0.99	0.99	1.00	0.99	0.99	1.00	0.99	0.99	0.66	0.54	0.54
$M = \{2, 2, 2\}$	1.00	0.97	0.97	1.00	0.96	0.96	1.00	0.96	0.96	0.72	0.53	0.53
$M = \{5, 5, 5\}$	0.99	0.88	0.88	0.98	0.84	0.84	0.97	0.83	0.83	0.84	0.63	0.63
$M = \{15, 15, 15\}$	0.98	0.80	0.80	0.94	0.77	0.77	0.91	0.77	0.77	1.00	1.00	1.00
$M = \{19, 10, 7\}$	0.95	0.78	0.78	0.92	0.75	0.76	0.89	0.74	0.74	0.98	0.97	0.96
$M = \{13, 10, 9\}$	0.98	0.81	0.81	0.96	0.79	0.79	0.94	0.79	0.79	0.99	0.98	0.98
$M = \{7, 10, 11\}$	0.97	0.81	0.81	0.95	0.78	0.78	0.93	0.77	0.78	0.97	0.94	0.94
$M = \{1, 10, 13\}$	0.99	0.84	0.84	0.98	0.80	0.80	0.96	0.80	0.80	0.97	0.95	0.94
$\mu_2 = 30$	0.94	0.76	0.75	0.89	0.75	0.75	0.86	0.75	0.75	1.00	1.00	1.00
$\mu_2 = 120$	1.00	0.89	0.89	0.99	0.85	0.85	0.98	0.83	0.83	0.90	0.64	0.64
$c_h = 0.25$	1.00	0.97	0.97	1.00	0.97	0.97	1.00	0.97	0.97	1.00	1.00	1.00
$c_h = 0.5$	1.00	0.92	0.92	1.00	0.91	0.91	1.00	0.91	0.91	1.00	1.00	1.00
$c_p = 30$	1.00	0.95	0.95	1.00	0.94	0.94	1.00	0.95	0.95	1.00	1.00	1.00
$c_p = 50$	1.00	0.92	0.92	1.00	0.91	0.91	1.00	0.91	0.91	1.00	1.00	1.00
$c_s = 5$	0.99	0.84	0.84	0.98	0.81	0.81	0.96	0.81	0.81	1.00	0.99	0.99
$c_s = 20$	0.87	0.75	0.74	0.84	0.71	0.71	0.82	0.71	0.71	0.78	1.00	1.00
$c_l = 0$	0.96	0.80	0.80	0.93	0.76	0.76	0.90	0.76	0.76	0.95	1.00	1.00
$c_l = 5$	0.98	0.81	0.81	0.95	0.79	0.79	0.93	0.77	0.77	0.98	1.00	1.00
Average	0.98	0.86	0.86	0.96	0.84	0.84	0.95	0.83	0.83	0.93	0.90	0.90

*$M = \{10, 10, 10\}$, $\mu_2 = 60$, $c_h = 1$, $c_p = 70$, $c_s = 10$, $c_l = 10$

online orders each store has to fulfil decreases. As a result, each store orders fewer products, increasing the probability of a stock-out. However, with more stores in the retailers network, the probability of a network-wide stock-out becomes lower as only one store needs a product for the cycle service level to be one for the online channel. Especially the MFP and TMFP have a high probability of stock-out when there are few stores in the network available, as under these policies already fewer products are ordered.

In the instances in which the number of store per type remains equal it was observed that more stores resulted in lower cycle service levels for stores. A similar trend is observed when altering the number of store per type while remaining the total average in-store demand equal. Under the OSIP the online channel is not much influenced by the different instances, but the total number of stores is quite high thus the probability of a network-wide stock-out is low.

Having a higher average online demand decreases the cycle service level, although the retailer will order more products to compensate for the higher demand the variance of demand is also higher. Compensating for this fluctuation with a safety stock can become to costly therefore the retailer prefers to incur lost sales.

Having a lower holding and procurement cost causes the retailer to stock more products, as it increases the profit margin of an individual product. Therefore the cycle service level increases in those instances. With an increase in shipping cost the profit margin of online sold products decreases thus the cycle service level also decreases as the retailer will be more willingly to incur a lost sale. With a higher penalty cost for cancelling an online order, the cycle service level increases as the retailer will stock more products to prevent incurring a high penalty cost.

Inventory level

As seen in Section 4.5.2, the inventory level under the OSIP fluctuates less than under the MFP and TMFP. To compare the fluctuation of the inventory under different online fulfilment policies we use the coefficient of variation (CV), i.e. the standard deviation of the store inventory level divided by the average store inventory level. We are only interested in the CV on the day before replenishment as a large CV on this day indicates a large probability of a stock-out. Table 4.5 presents the CV on the day before replenishment for all the different instances.

Table 4.5: Coefficient of variation for the inventory level before replenishment of all instances.

	Small store			Medium store			Large store		
	OSIP	MFP	TMFP	OSIP	MFP	TMFP	OSIP	MFP	TMFP
Base test case*	0.20	0.55	0.55	0.16	0.51	0.50	0.16	0.47	0.47
$M = \{1, 1, 1\}$	0.28	0.52	0.52	0.27	0.48	0.48	0.26	0.44	0.44
$M = \{2, 2, 2\}$	0.26	0.62	0.62	0.24	0.52	0.52	0.23	0.46	0.46
$M = \{5, 5, 5\}$	0.24	0.56	0.56	0.21	0.49	0.49	0.17	0.45	0.45
$M = \{15, 15, 15\}$	0.18	0.56	0.56	0.19	0.54	0.53	0.18	0.50	0.49
$M = \{19, 10, 7\}$	0.20	0.55	0.54	0.17	0.52	0.51	0.18	0.50	0.49
$M = \{13, 10, 9\}$	0.19	0.55	0.55	0.15	0.51	0.50	0.16	0.48	0.47
$M = \{7, 10, 11\}$	0.22	0.55	0.54	0.17	0.50	0.50	0.16	0.47	0.46
$M = \{1, 10, 13\}$	0.19	0.54	0.54	0.18	0.50	0.49	0.14	0.47	0.46
$\mu_2 = 30$	0.25	0.61	0.61	0.24	0.56	0.55	0.24	0.51	0.51
$\mu_2 = 120$	0.18	0.52	0.52	0.16	0.46	0.45	0.12	0.43	0.43
$c_h = 0.25$	0.10	0.44	0.44	0.10	0.40	0.40	0.10	0.36	0.36
$c_h = 0.5$	0.11	0.53	0.53	0.12	0.47	0.47	0.13	0.43	0.42
$c_p = 30$	0.11	0.50	0.50	0.11	0.44	0.44	0.11	0.39	0.39
$c_p = 50$	0.12	0.53	0.53	0.11	0.47	0.47	0.13	0.43	0.43
$c_s = 5$	0.17	0.56	0.56	0.16	0.51	0.51	0.14	0.47	0.47
$c_s = 20$	0.26	0.51	0.51	0.20	0.49	0.48	0.18	0.47	0.46
$c_l = 0$	0.20	0.53	0.53	0.18	0.51	0.50	0.16	0.48	0.47
$c_l = 5$	0.20	0.55	0.55	0.16	0.50	0.50	0.16	0.48	0.47
Average	0.19	0.54	0.54	0.17	0.49	0.49	0.16	0.46	0.45

*$M = \{10, 10, 10\}$, $\mu_2 = 60$, $c_h = 1$, $c_p = 70$, $c_s = 10$, $c_l = 10$

From Table 4.5 it is observed that the CV of the OSIP is more than half of that from the MFP and TMFP for the base test case, which is inline with the findings from Figure 4.4. Furthermore, it is observed that the CV of small stores is bigger than the one for large stores — which is expected as the demand for small stores also has a higher CV.

The CV of the OSIP differs between instances, but remains stable under the MFP and TMFP. Although the CV under the OSIP is more sensitive towards the studied settings, it remains lower than those of the MFP and TMFP. We observe that under the OSIP, with more stores in the retailers network, the CV drops. As more stores are available in the network the online demand can be fulfilled from more stores, therefore the retailer can more easily balance the store inventories in the network.

The instances in which the retailer orders more product, due to higher online demand or low holding, procurement, or shipping cost, the CV of the stores is also lower. Due to higher average inventory levels, while having equal CV of demand, the CV of the inventory becomes lower.

In summary, we observe that the OSIP outperforms both the MFP and TMFP under different retail network settings on all key performance indicators. The OSIP has the advantage of making better online fulfilment decisions, as it better balances direct and future expected rewards. The OSIP performs better when there are many stores in the retailers network or when the fraction of online demand compared to in-store demand is low. Especially the service level of both the offline channel and online channel is much higher for those instances. When the retailer orders more products for each store, the online fulfilment decision becomes less important thus the gap between OSIP and the two heuristic becomes closer to zero.

4.6 Conclusion and Discussion

Brick-and-mortar retailers are increasingly using their store inventory for the fulfilment of online orders. The decision of which store to use for the fulfilment of the online orders is however not straightforward, as the retailer has to take into account all store inventory levels, outstanding orders, and potential future demand. Therefore current practice is often to use myopic order allocation rules. However, by not taking into account future potential demand, inventory levels become imbalanced across the retail store network which can result in expensive excessive stocks for some stores while others might face stock-outs.

This paper contributes to the academic literature by providing a method to incorporate the online fulfilment decision and ordering decision for an omni-channel retailer network. Previous research does not consider a multi-period and dynamic order policy with non-zero lead times, as this can easily make the problem intractable with a large number of stores in an omni-channel retail network. To find a good performing policy for both the online fulfilment and ordering decision while still being able to solve the problem within reasonable time, we develop an approach in which we apply one-step policy improvement on a decomposed MDP. We solve the MDP for individual stores with value iteration, and use the relative value vector for the one-step policy improvement. This approach has the advantage of making a well-informed online fulfilment decision as it is able to capture the trade-off between fulfilling an online order or reserving the inventory for future demand. We compare the online fulfilment policy found by one-step policy improvement to two heuristics that are found in practice.

Based on an extensive numerical investigation, we can conclude that the one-step policy improvement outperforms both heuristics, on profit, cycle service level, and inventory balance. Although the profit gap between the one-step policy improvement and the two heuristics is only around 1.0%, the cycle service level is much higher for all stores and the online channel under the one-step improved policy. Furthermore, the one-step policy improvement is able to keep the store inventories from fluctuating too much, indicating that the inventory balance across the retailer's store network is more balanced, which is advantageous for the retailer as it results in fewer excessive stocks for some stores and fewer stock-outs for other stores. Moreover, when demand is fulfilled from stores, the store inventory needs to be increased. This has the added advantage that the service level for the in-store customer is increased. Also for the online channel higher service rates is achieved as demand is not limited to one stock, as now there are multiple locations available for fulfilling an online order.

In this paper, we investigated how we could make better informed online order allocation decisions by incorporating potential future demand using the one-step policy improvement. Although the model takes into account most characteristics of an omni-channel retail network, several potentially relevant factors were not included and which could be directions for further research. Firstly, the model does not take into account the location of the store and customer. Even though shipping rates are not always dependent on the distance (especially when these distances are relatively low), from an environmental perspective using a store to fulfil the demand of a nearby a customer is more sustainable and allows for other modes of transportation such as cargo bikes. Other modes of transport also have different cost structures, therefore the online fulfilment decision might become dependent on the shipping distance.

Secondly, we only consider customers who want to fulfil their demand via buying in-store or purchasing online. However, many customers are also interested in buying online, and then picking up their order in a store. Such a fulfilment option has different costs and fulfilment structure. Store pickup is discussed extensively in recent literature, but the related order allocation is an interesting research direction. Although pick-up orders are for a specific store, other stores could potentially also fulfil the order.

Thirdly, omni-channel retailers operating multiple stores and an online channel might also have an online fulfilment centre if their online demand is significant. Even in settings with an online fulfilment centre, store inventory might still be utilised to additionally fulfil online orders. Therefore an interesting research direction is to investigate under which circumstances the stores will still be used for the online fulfilment process. Additionally, the results of the introduction of an online fulfilment center could be used to compare settings without an online fulfilment center. This could provide valuable managerial insights about the economic added value of introducing such a fulfilment location.

Fourthly, we do not consider the setting where the retailer can use lateral transhipment to redistribute the stores' inventories. This might be of interest to the retailer is that it can reduce overstocking in one store and understocking in the other. Lastly, the retailer might want to protect some of the store inventory for in-store consumers and thus determines that a minimum amount of inventory that should always be in-store. This could improve service levels for in-store consumers, such a decision is interesting for future research.

Lastly, new retail concepts such as instacart and darkstores also rely on the fulfilment of online orders from physical locations. With these concepts, online orders often can be fulfilled from multiple locations within the city. The allocation decision does need to be made in a shorter time- frame as these services provide fast delivery times. Nonetheless, these concepts fit in the domain of ship-from-stores and could be an interesting research direction.

Chapter 5

Inventory and Fulfilment Decisions for Different Omni-channel Concepts in a Retail Network Setting

This chapter is published as:

Goedhart, J., Haijema, R., Akkerman, R., & S. de Leeuw (*Under review*). Inventory and fulfilment decisions for different omni-channel concepts in a retail network setting

Abstract

Omni-channel retailing allows customers to fulfil their demand from anywhere, online or offline, and receive the product in their preferred manner, be it at home or in the store. Orders that are placed online can be fulfilled from the inventory of a store or an Online Fulfilment Centre (OFC). For an omni-channel retailer, deciding on the type of network setting to fulfil the different types of demands can be challenging, as different trade-offs exist between inventory costs, order fulfilment costs, and service levels for different sales channels. The potential synergies resulting from combining different fulfilment options are also not straightforward. In this paper, we therefore investigate four typical network settings used in practice with a focus on the inventory management of the stores and OFC. To analyse this process, we model the fulfilment and ordering decisions as a Markov Decision Process that maximises the expected profit, which we solve via a one-step policy improvement approach. Based on an extensive numerical study, we show that allowing both stores and OFC to fulfil demand is most profitable, but that only using stores would result in higher overall fill rates. Our results also show the added value of using the store or OFC for the fulfilment of orders placed online, and how this depends on the inventory cost and order fulfilment cost differences between the locations. With minimal cost differences, the results indicate that including stores in the fulfilment process is preferred.

5.1 Introduction

To compete with online retailers, brick-and-mortar retailers are adopting online channels to integrate with their current channels (McKinsey, 2021c). The integration of online and offline shopping channels is called omni-channel retailing and is becoming an increasingly important strategy for brick-and-mortar stores to compete with pure-online retailers (Melacini et al., 2018). Fully integrating the different shopping channels allows customers to fulfil their demand from anywhere and receive the product in their preferred manner, be it in the store, at a designated pickup point, or at home. The core of omni-channel retail is in the possibility for customers to seamlessly switch between channels during a purchasing process (Hübner et al., 2016a). Due to these developments towards omni-channel retail strategies, the role of the store in retail networks needs reconsideration. A store is not just a place to shop and buy, it is also an asset that can support the online channel (Hübner et al., 2022).

Omni-channel retailing allows retailers to fulfil demand in different ways, even though each different type of demand has its specific needs. The customers demand is taking place either online or offline, where the customer places an order via a web-page for example or by visiting a store. When buying through the online channel, a customer can often choose different ways to get the ordered product: they can have it shipped to their home or collect in a store. The former we call buy-online-deliver-at-home (BODH) and the latter buy-online-pick-up-in-store (BOPS). Therefore we can identify three different types of customers in omni-channel retailing: walk-in, BOPS, and BODH. There is also the case in which a walk-in customer orders a product in-store to be delivered at home, but as this order needs the same handling as product ordered online we can consider them as an BODH customer. This variety in customer types increases logistics complexity and retailers therefore need to strategically reconsider their store network, as these stores can support the online channel (Baird & Kilcourse, 2011).

The retailer needs to decide how to fulfil the demand of the customer. The retailer has little control over walk-in customer, but for products ordered online, it has different options for the fulfilment. If the customer wants to have their online order shipped to their home the retailer can decide to ship the order from a store or from an online fulfilment centre (referred to as OFC). Shipping online orders from the store to the customer is referred to as a ship-from-store (SfS) strategy, which is the only option for retailers that do not have a dedicated OFC in their network (Agatz et al., 2008).

However, retailers with an OFC can still choose to use stores for fulfilment of online orders as this can be more cost-efficient.

For the customers that ordered their product online but want to collect it in-store the retailer has a similar decision to make. Such a BOPS order can be fulfilled from in-store inventory or the order can be fulfilled by the OFC. If the OFC fulfils the BOPS order, the product will be picked and packed at the OFC and then shipped to the store for pick-up. This strategy is referred to as a ship-to-store strategy (Akturk et al., 2018).

Using stores for online fulfilment has low set-up costs and is thus attractive for retailers that have just started an online channel. An additional advantage for using stores for online fulfilment is that a demand pooling effect may be created when store inventory is used for both walk-in and online customers, which can decrease out-of-stock probabilities and thereby increase sales (Bayram & Cesaret, 2021).

However, using stores for the fulfilment of online orders also has disadvantages, as stores are often not designed for order picking and packing, and thus related operational costs can be high (Ishfaq & Raja, 2018). Therefore, retailers with a high volume of online sales may prefer to use an OFC for the fulfilment of online orders as these are designed for efficient picking, packing, and shipping online orders (Hübner et al., 2016a). The advantage of this is that all online demand can be pooled and that inventory costs at an OFC are often lower than in stores. However, this would require coordination with stores already holding inventory, which could also be used for BOPS orders. Therefore, deciding the role of an OFC in the fulfilment of online orders is a tactical network decision that is important as it influences the role of the store and the cost structure of fulfilling BOPS and BODH orders.

Even though fulfilling multiple types of demand from a single location (store or OFC) is likely to lead to inventory pooling effects, it can be hard to have these effects materialise in practice in omni-channel networks with a range of fulfilment options. Some demands are region-dependent as walk-in and BOPS customers may only be willing to travel to a local store. However, BOPS demand can also be fulfilled via an OFC, which has the advantage that all BOPS orders can be pooled at a central place and picking and packing is more efficient due to higher volumes (Hübner et al., 2016a). For BODH customers preferring home delivery, any store or OFC is eligible to fulfil their demand. Thus, for the fulfilment of walk-in and BOPS customers, it is important that the demand can be fulfilled in the region, whereas for the BODH orders products need to be available in the retailers network. In the design and operation of their network, an omni-channel retailer therefore needs to determine what the best fulfilment strategy is for their situation. Furthermore, a demand fulfilment decision needs to be made to identify from which location to fulfil an actual order. An omni-channel retailer needs to take into account several complicating factors such as demand and inventory pooling, picking, packing and shipping, cost savings, and product availability when making these decisions.

Hübner et al. (2022) describe common omni-channel retail fulfilment strategies that we use as a starting point. Figure 5.1 depicts the four fulfilment strategies. The rectangles depict the fulfilment locations and the ellipses the customer fulfilment options; the arrows identify which fulfilment options are viable for which fulfilment locations.

Figure 5.1(a) depicts the *store fulfilment* (SF) strategy (or SfS) in which only stores are used as fulfilment locations. Therefore if a customers want to fulfil their demand, either via walk-in, BOPS, or BODH, it will be fulfilled from store inventory. Most retailers who start an online channel, first leverage their store network to fulfil online orders. The advantage of this situation is that demand of all different channels is pooled and therefore could reduce overstocking costs at the stores. Furthermore, online orders can be assigned to different stores, allowing the retailer to prevent imbalances across the store network. Disadvantages are that stores are relatively expensive to fulfil online orders due to higher holding, picking and packing costs as they are not designed for picking.

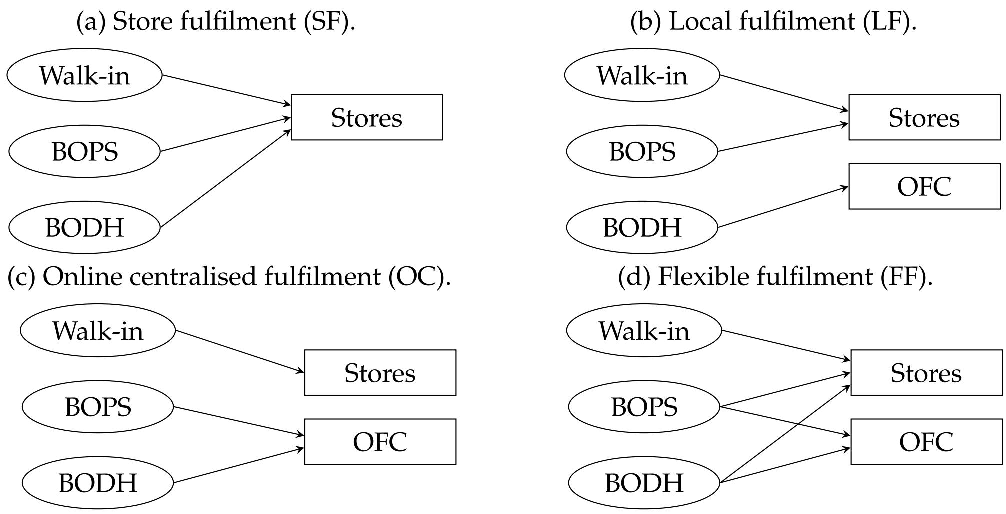


Figure 5.1: The different fulfilment strategies of an omni-channel retailer.

When introducing an OFC for the online channel, the decision needs to be made if the OFC is used as fulfilment location for the BOPS demand. Figure 5.1(b) depicts the *local fulfilment* (LF) strategy in which the OFC is a fulfilment location for all BODH demand and the stores are fulfilment locations for walk-in and BOPS demand. Thus store inventory is used for the customers who are willing to fulfil their demand by visiting the store, albeit by ordering online or walk-in whereas customer wanting the product to be shipped to their home the OFC inventory is used. The advantage of this strategy is that the BOPS order do not need to be shipped from an OFC to the store, however the disadvantages is that demand is only pooled at the store.

Figure 5.1(c) depicts the *online centralised fulfilment* (OC) strategy in which the OFC is used as fulfilment location for BOPS and BODH demand. This means that BOPS orders are shipped from the OFC inventory to the store (referred to as ship-to-store) to be picked up. An advantage of using an OFC for BOPS demand is pooling of inventory for meeting the BOPS demand of all stores at one central place. Furthermore, holding costs are typically lower at an OFC. The disadvantage of fulfilling a BOPS order from the OFC is that there is a shipping cost as the product needs to be shipped from the OFC to the store.

Figure 5.1(d) shows *flexible fulfilment* (FF) in which customer who place their order online can be fulfilled by any fulfilment location, stores or OFC. Although flexible demand fulfilment might be the ideal solution as it has all the pros of the other fulfilment strategies without the cons, it needs the coordination of sales channels, stores, and an OFC to be seamless. It is however difficult to assess the added value of flexible fulfilment compared to other strategies in order to evaluate if it outweighs the increase in complexity of channel coordination and costs of installing an OFC.

Since demand can be fulfilled from multiple locations and inventory is (partially) pooled across the network, it is difficult to assess which fulfilment strategy is the best for a specific retailer and customer. The retailer needs to decide which fulfilment location to use for which type of demand, with each possibility having different costs. Additionally, the effect of pooling certain types of demand at a fulfilment location is not straight-forward and difficult to asses. Furthermore, it is not clear whether enabling different fulfilment options leads to sufficient benefits to warrant a possible increase in complexity. In this study, we therefore investigate the different fulfilment strategies and their effect on profitability and service levels.

Only few papers study omni-channel fulfilment strategies in the context of a retail network with multiple stores. Such problems require the simultaneous decision-making of the fulfilment strategy and from which inventory to fulfil demand (Hübner et al., 2022). As retailers are adopting more fulfilment options in their retail practice, this is an interesting research direction. Most of the existing literature focuses either on the strategic decision of what type of omni-channel concept to adopt (e.g. Bayram & Cesaret, 2021; Saha & Bhattacharya, 2021) or on the operational decision of the demand fulfilment (e.g. Andrews et al., 2019; Govindarajan et al., 2021). However, only few studies incorporate both these decisions for an omni-channel retailers network.

In this paper, we study the replenishment and demand fulfilment decisions of an omni-channel retailer that operates multiple brick-and-mortar stores and an OFC. We investigate multiple different fulfilment strategies, in which the stores and an OFC are involved differently in fulfilling BOPS and BODH demand. For each of the stores, the retailer must decide how many products they want to make available for the BOPS and BODH channel, considering the different cost structures of each fulfilment option and the expected sales. The retailer has to allocate all received BOPS and BODH orders to either a store or an OFC depending on the fulfilment strategy chosen. This decision depends on the involved fulfilment costs as well as the inventory positions of the stores. Fulfilling an order from store inventory could results in a stock-out experience to an walk-in customer.

With this study, we contribute to the scientific literature in several ways. We research multiple omni-channel fulfilment strategies, explicitly taking into account the operational inventory management decisions (i.e. replenishment and demand fulfilment). Based on our numerical results we show that only using store inventory for the fulfilment of demand results in a lower profit but higher fill rates. Furthermore, only using an OFC for BOPS and BODH demands results in relatively low profits and fill rates. Finally, we show how the choice of fulfilment strategy depends on the relative differences between cost factors, where the differences between inventory and fulfilment costs at store level and at OFC level determine how store fulfilment and OFC fulfilment are used in the best possible way.

The remainder of this paper is structured as follows. Section 5.2 presents related research on fulfilment strategies for omni-channel retailers. In Section 5.3, we describe the decision problem and the formulation as an MDP. In Section 5.4, we investigate the results of the different fulfilment strategies on several instances and derive insights. Section 5.5 concludes the paper and discusses future research directions.

5.2 Literature Review

As described in Hübner et al. (2022) the decisions related to omni-channel retail network settings can be classified into two planning horizons: the fulfilment strategy and the demand fulfilment problem. Therefore in the literature one can identify two different literature streams, we first briefly describe literature related to the strategic decision which considers whether to implement an online channel and which locations to be used in the fulfilment. In the second part we will focus on literature that investigates the demand fulfilment decision for a retailer that has one or more stores available to fulfil the demand of the online channel, as we are more interested in the operational decision-making. In the last section we focus on literature that considers both the strategic and operational decision for an omni-channel retailers network.

5.2.1 Strategic decisions - channel portfolio and fulfilment locations

Literature related to the strategic decision is focused on whether to open an online channel, whether to include BOPS or which locations should be used to handle the different types of demand. Ishfaq & Raja (2018) investigates whether to use dedicated OFC for the fulfilment of online orders or to leverage the stores assets for the online orders. Via an empirical study they compare the different fulfilment options, they conclude that the operational, picking and packing costs greatly influence the fulfilment option. Similarly, Li et al. (2022) also researches whether to use an SfS strategy or an OFC for fulfilling online orders, while taking into account the risk behaviour of the retailer. They conclude that if the inventory cost of an OFC is high or the retailer is less afraid of risk, fulfilling orders with store inventory is preferred.

Saha & Bhattacharya (2021) researches under which situations store inventory should be used for the fulfilment of BOPS orders. When the profit for the store retailers for BOPS orders or quantity of BOPS orders is low the store will reject BOPS orders and prioritise walk-in customers. Hu et al. (2022) also investigates the influence of implementing BOPS, concluding that when walk-in consumers migrate to purchasing via BOPS it can result in lower store inventories and more stock-outs due to lower profit margins under BOPS. However, if the retailer can induce online consumers to switch to BOPS it results in demand pooling at the store and can benefit the retailer.

However, the strategic literature only considers either BOPS or SfS as a fulfilment option but does not consider both. Furthermore, they do not consider an online fulfilment decision as this is an operational decision. However, the online fulfilment decision might influence which fulfilment strategy is most profitable and therefore should be included in the analysis. Thus the retailer has to make a strategic decision of which omni-channel concept to be used and an operational how to manage the inventory (Hübner et al., 2022).

5.2.2 Operational decisions - order allocation

The operational problem is related to the fulfilment of the demand of the online channel to the locations. One of the first papers to research the demand fulfilment decision is by Xiao et al. (2009). They investigate for a two store retail network with a limited selling horizon from which store to fulfil online demand. Bretthauer et al. (2010) and Mahar et al. (2009) research a similar setting though with more stores. The latter paper improves the policy of Bretthauer et al. (2010) by taking into current inventory positions which reduces overall costs and increases online sales. However, for large problem instances deriving a policy becomes too complex and makes the problem intractable. Ni et al. (2019) focus on making the demand fulfilment decision for same-day delivery via crowdshipping. They develop an algorithm that makes the demand fulfilment decision based on minimising delivery cost and expected future cost. Their approach outperforms common retail strategies for the demand fulfilment decision. Andrews et al. (2019) develops an algorithm for the fulfilment of demand to either stores or an OFC based on available inventory, delivery time, and costs. Zhou et al. (2021) researches a similar setting but also includes a pricing decision in their problem. Both studies do not include a replenishment decision in their model and assume that an OFC has enough stock to fulfil any potential order if stores have no available inventory. Govindarajan et al. (2021) does include a replenishment decision in their model that also makes a demand fulfilment decision, however they decouple the ordering and fulfilment decision to keep the model tractable. They discuss that the added value of using store inventory for fulfilment is the synergy of pooling walk-in and online demand within and across locations. They conclude that including BOPS in the model is an important research direction. Also, Jiu (2022) does include a multi-period replenishment decision in their study about how SfS can improve omni-channel retailers, it also does not include BOPS demand. Pichka et al. (2022) investigates the interesting case where in-store demand can also be fulfilled via an OFC, however stores are still included in the online fulfilment decision. They do not take into account a replenishment decision as the current decision-making is already too complex.

5.2.3 Integration of strategic and operational decisions

Arslan et al. (2021) is the only paper that researches the demand fulfilment decision for different fulfilment strategies. They consider three different locations (OFC, stores, urban fulfilment platforms) from which demand can be fulfilled. Their approach consists of two stages, where at the first stage the fulfilment locations are selected to fulfil a set of online demand and in the second stage it fulfils the demand from different locations. They discover that using stores for the fulfilment of online demand can increase profitability, mainly due to more online demand being satisfied. They mention that an interesting research direction is to include other omni-channel retailing fulfilment options such as BOPS. Lin et al. (2022) integrates the strategic and operational decision by deciding which stores should be included in the fulfilment decision and then also how much of the online sales the store should fulfil. They construct a heuristic to solve the problem, which outperforms other benchmarks.

Only limited papers consider both the BOPS and SfS strategy as omni-channel strategy. Wang et al. (2022) compares StS and SfS, where they investigate the effect of implementing StS or SfS. Their study is limited to a setting of only one store and an OFC and only focused on the replenishment decision. They conclude that implementing StS or SfS results in lower total order quantities as excess demand can now be fulfilled from multiple locations. Yang et al. (2022) researches a similar setting but includes a rationing decision determining the quantity of inventory available for each channel. They similarly mention that introducing BOPS or SfS influences the replenishment decision, but argue that it also depends on related cost factors such as picking cost. They mention that for large retailers a fully integrated omni-channel retail strategy can be profitable, while for smaller retailers only introducing BOPS or SfS is more cost-effective.

5.2.4 Research gap

From the literature it is seen that no paper combines the decision on which fulfilment location to use for different types of demand with the decision of the fulfilment of demand. More specifically, no paper investigates the (dis-)advantages of different fulfilment strategies for the three types of demand of an omni-channel retail network while taking the inventory decision such as the replenishment and demand fulfilment decision. However, as retailers are adopting more fulfilment options in their retail practice this decision becomes a more pressing issue. Therefore in this study, we investigate an omni-channel retailer who operates multiple omni-channel concepts in a network setting of multiple stores and who tries to optimise the total expected profit.

5.3 Methodology

We will compare the four fulfilment strategies for an omni-channel retailer with J stores and an optional OFC. We first describe the full fulfilment strategy in which both the stores and OFC can fulfil BOPS and BODH orders (Figure 5.1(d)), from this model the other fulfilment strategies can be derived as explained in Appendix 5.A.

5.3.1 Optimisation problem

The set of physical stores are indexed $j \in \{1, 2, \ldots, J\}$, and index $j = 0$ represents the OFC. The set $\mathcal{J} \in \{0, 1, \ldots, J\}$, is the set of all $J + 1$ locations, from which demand can be fulfilled. The retailer has three different types of demand: walk-in, BOPS, and BODH. Here, walk-in demand relates to customers who buy and collect a product upon visiting a physical store. BOPS orders are orders placed online by customers who want to pick up their product at their preferred store. Therefore the order can be either fulfilled from the preferred store inventory or by a product sent from an OFC to the store. BODH orders are to be shipped to a customers home directly from either a store or an OFC. The fulfillment costs (consisting of shipping, picking and packing costs) of a BOPS or BODH order depends on whether it is fulfilled by the store or OFC.

The retailer periodically sets a replenishment order and R subsequent demand fulfilment decisions, as depicted in Figure 5.2. A period is thus divided into R sub-periods. At the start of every period the retailer determines the order quantity (q_j) for the J stores and for the OFC (q_0). The replenishment happens $L \leq R$ sub-periods later. During each sub-period uncertain demand accumulates and is fulfilled at the end of each sub-period. If $R = 7$ this could reflect a weekly replenishment decision, and a daily demand fulfilment decision.

The three sales channels are indicated by index $i \in \{1 = \text{Walk-in}, 2 = \text{BOPS}, 3 =$

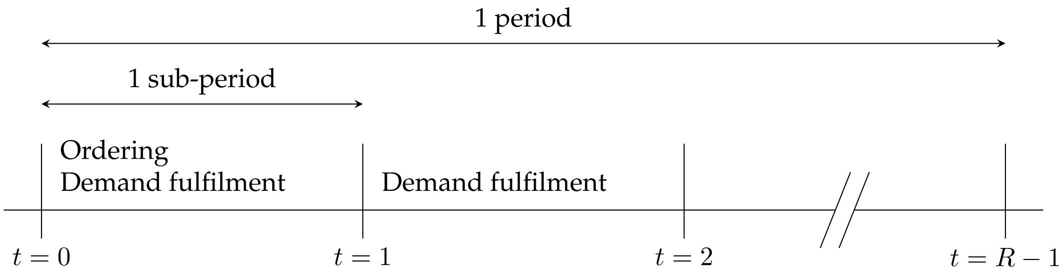


Figure 5.2: Visualisation of a period, sub-period and events: replenishment orders are placed every R sub-periods, whereas demand fulfilment allocation happens at the end of every sub-period.

BODH}. Let $d_{1,j}$ be the quantity demanded during a sub-period by walk-in customers at store j, and $d_{2,j}$ be the quantity that customers have ordered online and wish to pick at store $j \neq 0$. The BODH demand during a sub-period is denoted d_0. Before the BOPS and BODH demand is fulfilled, first the walk-in demand $d_{1,j}$ is met from the remaining stock at store j, as walk-in customers immediately take the products with them. Next, BOPS orders are fulfilled from either the respective store, or from the OFC. As BOPS demand cannot be fulfilled by other stores, fulfilling BOPS demand gets priority over BODH demand. Finally, the BODH demand is fulfilled by considering the remaining stock at all stores and at the OFC.

The replenishment and demand fulfilment decisions are made periodically such that the long-run average or discounted profit is maximised. Profit equals the sales revenue minus the procurement cost, holding costs and costs for shipping, picking and packing product. As common in omni-channel retailing the sales price is equal in all channels, but holding and picking and packing costs may depend on where products are stored and handled. In case of a lost sale, a unit shortage penalty applies on top of the implied lost profit margin.

The sequential periodic decision problem is modelled as an infinite horizon Markov Decision Process (MDP). In the next subsection, the main components of the MDP model are presented. A more detailed description is found in the appendix. In Subsection 5.3.3, a one-step policy improvement approach is presented to (approximately) solve the problem.

5.3.2 MDP formulation

States, actions, and state transitions

We formulate the problem as a Markov Decision Process (MDP), with state $s = (t, e, I, Q, B, A)$, where:

- $t \in \{0, 1, \ldots, R-1\}$ = index of sub-period,
- $e \in \{1 = \text{Walk-in}, 2 = \text{BOPS}, 3 = \text{BODH}\}$ = the decision epoch within a sub-period, which relates to the order in which the different demands are fulfilled,
- $I = (I_0, I_1, \ldots, I_J)$ = the inventory position of each store and OFC at the start of epoch e,
- $Q = (Q_0, Q_1, \ldots, Q_J)$ = the outstanding orders of the stores and OFC at the start of epoch e,
- $B = (B_1, B_2, \ldots, B_J)$ = for each store, the BOPS demand to be fulfilled in sub-period t,
- A = the BODH demand to be fulfilled in sub-period t.

Without loss of generality, we assume replenishment orders are set at the start of the sub-period $t = 0$, and are delivered at the start of sub-period $t = L$. At the end of each sub period the demand accumulated at the three different sales channels will be fulfilled. We model this integrated decision as three sequential decisions. Therefore we split the sub-period into three decision epochs, as depicted in Figure 5.3.

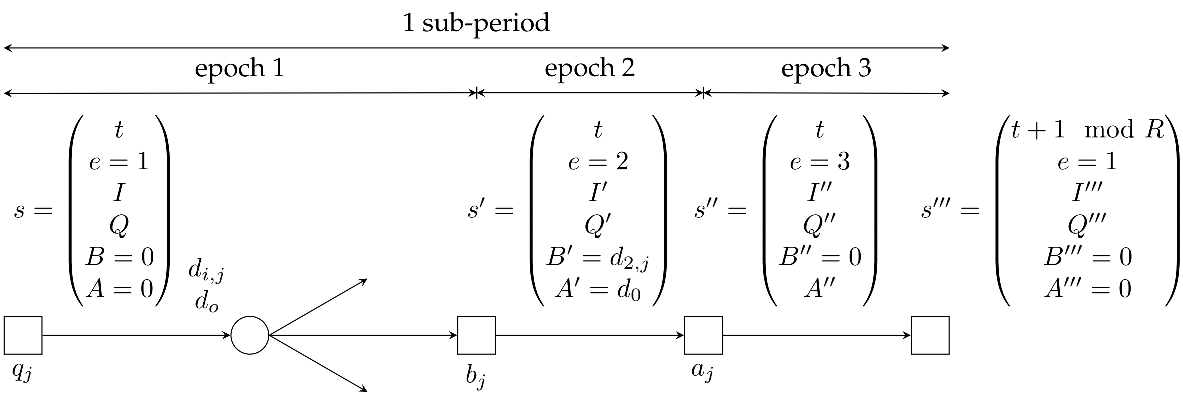


Figure 5.3: Visualisation of the chronological order of the states, actions, and stochastic processes.

The chronological order of actions and events is presented in Figure 5.3. A square symbol indicates a state where a decision is to be taken, and a circle indicates a stochastic process. At the beginning of sub-period $t = 0$, the order decision q_j is to be made for each store and the OFC, based on the actual stock levels $I = (I_0, I_1, \ldots, I_J)$. For $t \in \{0, 1, \ldots, R-1\}$, this decision is skipped. As the quantity ordered should be tracked for L sub-periods, any outstanding orders are included in the state transition, i.e.:

$$Q' = \begin{cases} (q_0, \ldots, q_J) & \text{if } t = 0 \\ (0, \ldots, 0) & \text{if } t = L \\ Q & \text{otherwise.} \end{cases}$$

The uncertain demand during a sub-period is modelled in epoch 1 by finite discrete probability distributions $P_{1,j}$, $P_{2,j}$, and P_0 for respectively the walk-in demand, the BOPS demand, and the BODH demand. The demand of all probability distributions are modelled by Poisson distributions with as average $\mu_{1,j}$, $\mu_{2,j}$, and μ_0 respectively. The distribution functions are truncated at maximum demand levels $D_{1,j}$, $D_{2,j}$, and D_0 respectively.

At the end of epoch $e = 1$, the demand at each of the three sales channels $(d_0, d_{1,j}, d_{2,j})$ becomes known and the inventory levels are adjusted to I': $\forall j \in \{1, 2, \ldots, J\} : I'_j = (I_j - d_{1,j} + \delta(t = L) \cdot Q)^+$ where $x^+ = \max(0, x)$, $\delta(x)$ denotes the Kronecker delta, which returns the value 1 if $x =$ True, otherwise 0, and $I'_0 = I_0$. The BOPS demand $d_{2,j}$ and the BODH demand d_0 are to be fulfilled in subsequent epochs, and are thus added to the state space: $B' = (d_{2,1}, \ldots, d_{2,J})$ and $A' = d_0$.

At the start of epoch $e = 2$, based on state $s' = (t, e = 2, I', Q', B', A')$, the retailer decides on how to fulfil for each store j the BOPS demand B_j. For each store j one decides on the quantity b_j to fulfil from the stock available at store j, and on the number of products b_0 to ship from the OFC to the stores to meet the residual BOPS demand: $\sum_{j=1}^{J} B_j - b_j$.

Thereby one may allow to not fulfil all demand, i.e. $b_j + b_0 \leq B_j$. This further reduces the available stock at the stores and at the OFC: $I''_j = I'_j - b_j$ and $I''_0 = I'_0 - b_0$. As the BOPS demand is not to be dealt with in epoch 3, it is excluded from the state space in epoch 3 by setting $B'' = 0$.

In epoch $e = 3$, the retailer decides on the fulfilment of the BODH demand d_0, based on $s'' = (t, e = 3, I'', Q'' = Q', B'' = 0, A'' = A'$. That is, one decides on the quantities $a = (a_0, a_1, \ldots, a_J)$ to deliver from the OFC and stores to meet the demand d_0. Any demand that is not fulfilled is considered to be lost thus $A''' = 0$ (which is in line with (e.g., Bayram & Cesaret, 2021)). The state s''' at the end of epoch 3, is the starting state of the next sub-period: $s''' = (t+1 \mod R, e = 1, I''' = I'' - a, Q''' = Q'', B''' = 0, A''' = 0)$. As the outstanding order state does not change during the second epoch either, $Q''' = Q''$. The BOPS demand state remains zero as no new orders are received during the epoch. Additionally, all BODH demand that could not be fulfilled is considered to be lost thus $A''' = 0$ which is a common assumption (e.g., Bayram & Cesaret, 2021).

Rewards and the Bellman equation

The aim of the MDP is to maximise the long-term weekly expected profit. The exact formulas of the expected revenue and costs per epoch are presented in Appendix 5.B. For sake of readability, we discuss the cost components and leave the notational burden. The expected profit in epoch 1, denoted as $C_1(s, q)$, is determined by the initial state s and the order decisions q, the demand probability distribution of the walk-in demand, the unit sales price p, and the unit purchase cost c. Unit holding cost are accounted over all products in stock at the start of a sub-period (in epoch 1 only). The unit holding cost h_j may depend on the location (OFC or store).

Lost sales in epoch 1 are not penalised. In epoch 2 and 3, however, not fulfilling online accepted orders is penalised by a unit shortage penalty cost g for each item ordered online that is not delivered. In addition, in epoch 2 and 3 we account for demand fulfilment costs, which consists of location-independent unit shipping cost k, and location-dependent unit product handling cost u_j. Product handling includes activities such as picking and packing, which may be less costly at the OFC than at a store, as an OFC is designed for such operations.

The expected profit in epoch 2, $C_2(s', b)$, is determined by the volume of the BOPS demand, the fulfilment decisions b, the unit sales price p, the demand fulfilment costs, and the unit penalty cost for each unit of BOPS demand that is not fulfilled. Similarly, the expected profit in epoch 3, $C_3(s'', a)$, is determined by the volume of the BODH demand, the fulfilment decisions a, the unit sales price p, the demand fulfilment costs, and the unit penalty cost for each unit of BODH demand that is not fulfilled.

Thus the expected profit in a sub-period is the sum of the reward functions C_1, C_2, and

C_3. The expected profit over n sub-periods when starting in state s and taking optimal decisions in this and future sub-periods is denoted by $v_n(s)$. The long-run expected profit over R successive sub-periods is $g = \lim_{n\to\infty} (v_n - v_{n-R})$. As the underlying Markov chain is ergodic, this difference converges to a constant g, which is independent of the initial state s.

To determine the value of g by a value iteration algorithm, one starts setting $v_0(s) = 0$ for all states s. The value of $v_n(s)$ for increasing values of n can be computed from equations (5.1) to (5.3):

$$v_n(s) = \max_{q \in \mathcal{Q}(t,I)} \left\{ C_1(s,q) + \sum_{j=1}^{J} \sum_{d_{1,j}=0}^{D_{1,j}} \sum_{d_{2,j}=0}^{D_{2,j}} \sum_{d_0=0}^{D_0} P_{1,j}(d_{1,j})\, P_{2,j}(d_{2,j})\, P_0(d_0) \cdot v'_n(s') \right\}, \tag{5.1}$$

where $v'_n(s')$ is computed from

$$v'_n(s') = \max_{b \in \mathcal{B}(I',B')} \left\{ C_2(s',b) + v''_n(s'') \right\}, \tag{5.2}$$

and $v''_n(s'')$ follows from

$$v''_n(s'') = \max_{a \in \mathcal{A}(I'',A'')} \left\{ C_3(s'',a) + v_{n-1}(s''') \right\}. \tag{5.3}$$

The iteration counter n is related to the number of sub-period over which the expected profit is maximised. This iterative procedure is called value iteration and may be stopped when the difference between v_n and v_{n-R} converges to a constant vector, i.e. stop if the span between v_n and v_{n-R} is smaller than a predefined accuracy ϵ. Here, the span is defined as $\|v_{n,j} - v_{n-R,j}\|$, which is the difference between the largest and the smallest element of $v_{n,j} - v_{n-R,j}$.

5.3.3 One-step policy improvement

The possible number of states $s = (t, e, (I_0, I_1, \ldots, I_J), (Q_0, Q_1, \ldots, Q_J), (B_1, \ldots, B_J), A)$ grows exponentially in the number of stores J. Computing an optimal policy is thus intractable in most cases with $J > 2$. This is called the curse of dimensionality. For an overview of the state and action space we refer to Appendices 5.C and 5.D. For the settings we wish to investigate, the problem cannot be solved within reasonable computation time by value iteration. Nevertheless, the MDP framework can be used to derive an approximate solution by a so-called one-step policy improvement algorithm. The principle is, first, to approximate the high-dimensional value vector by state values of a policy that is simpler to evaluate, and next execute one iteration step of the Bellman equation.

Decomposition of MDP

A common approach to simplify the evaluation of the state values, is to decompose the problem into MDPs with a lower state dimension. This decomposition approach has shown to perform well in variety of problem settings (Wijngaard, 1979; Bhulai, 2009; Haijema & Hendrix, 2014; Bhatnagar & Lin, 2019). We base our one-step policy improvement algorithm on the algorithm presented in Goedhart et al. (2023). The approach is extended by considering a network with an OFC and an additional fulfilment decision for the BOPS demand.

In our setting, the curse of dimensionality of the state space is circumvented, by decomposing the MDP, into $J + 1$ smaller MDPs: one for each of the J separate stores, and one for the OFC. Thus, to allow for such a decomposition, we implicitly model the effective demand, or fulfilment from the stores and from the OFC as independent processes, whereas in fact these are influenced by the demand fulfilment decisions. In the policy improvement step, we correct for this assumption by relaxing this initial allocation, and including the demand fulfilment decision.

The initial allocation of the mean BOPS and BODH demand to the stores and OFC is presented in Table 5.1. The allocation depends on the fulfilment strategy. For the fulfilment strategy (FF), we assume half of the BOPS demand is met from the stock at the store, the other half from the stock at the OFC, and all BODH demand will be shipped from the OFC. For network setting SF we approximate the number of BODH demand each store receives by evenly distributing the demand over all stores. For network setting FF we assume all BODH orders are fulfilled by the OFC, as fulfilling BODH orders via the OFC is less costly. For the BOPS orders we split the total evenly between the stores and OFC.

Table 5.1: Approximated average BOPS and BODH demand for the decomposed MDP for store fulfilment (SF), local fulfilment (LF), online centralised fulfilment (OC), and flexible fulfilment (FF).

Configuration		SF	LF	OC	FF
Stores	BOPS	$\mu_{2,j}$	$\mu_{2,j}$	0	$\frac{1}{2}\mu_{2,j}$
	BODH	μ_0/J	0	0	0
OFC	BOPS	0	0	$\sum_{j=1}^{J}\mu_{2,j}$	$\frac{1}{2}\sum_{j=1}^{J}\mu_{2,j}$
	BODH	0	μ_0	μ_0	μ_0

The relative value matrix of the problem for a single location $j \in \{0, 1, \ldots, J\}$ is denoted $v_{n,j}$. Theses states values can be computed for each state $s_j = (t, e, I_j, Q_j, B_j, A_j)$, using a lower dimensional version of equations (5.1) to (5.3). Thereby the probability distributions $P_0, P_{1,j}$, and $P_{2,j}$ are replaced by truncated Poisson distribution fitted to the means given in Table 5.1: $\widetilde{P}_0, \widetilde{P}_{1,j}$, and $\widetilde{P}_{2,j}$. When evaluating the state values for location j, the states space, action space, and demands related to all other stores can be neglected (e.g. set to zero).

With these assumption, we can solve the decomposed MDP for each store and OFC to find a relative value matrix to base a one-step improved policy on. The relative value matrix is found via value iteration where we maximise the expected weekly profit, where the problem is solved iteratively backwards using the Bellman equation as described in Subsection 5.3.2 for a single location. To ensure that all $v_{n,j}(s_j)$ are computed with the same number of iterations n, n is set to a value n_{max} to ensure that the MDP of each individual store converges to a span below $\varepsilon = 0.1$.

An algorithmic description of value iteration for the stores is given in Appendix 5.E, Algorithm 6. We set the demand distributions to the average demand given in Table 5.1. From the algorithm we receive the relative value matrices $v_{n_{max},j}$ of store j. The relative value matrix of the OFC $v_{n_{max},0}$ is received from Appendix 5.F, Algorithm 7.

The one-step-improved policy

For the decomposed problem we assume a fixed allocation of the (mean) demand over the stock locations. For the original not-decomposed problem, we compute a policy which improves this allocation. An approximately optimal action to take in a state is found by executing a single iteration of the Bellman equation with terminal reward set by $v_{n_{max},j}$. As we execute only one iteration, the new obtained demand fulfilment policy is called a one-step-improved policy. For simplicity we do not improve the order policy. For the ordering decision, we use the ordering policy of the decomposed MDPs.

The BOPS fulfilment decision using the one-step policy improvement is calculated as follow:

$$\widetilde{\pi}^b(s') = \arg\max_{b \in \mathcal{B}} \left\{ C_2(s', b) + \sum_{j \in \mathcal{J}} v_{n_{max},j}\left(s''_j\right) \right\}, \tag{5.4}$$

where the last term is the approximated expected profit when in state s'' derived from the relative value matrices.

The BODH fulfilment decision follows from:

$$\widetilde{\pi}^a(s'') = \arg\max_{a \in \mathcal{A}} \left\{ C_3(s'', a) + \sum_{j \in \mathcal{J}} v_{n_{max},j}\left(s'''_j\right) \right\}. \tag{5.5}$$

To evaluate the performance of the one-step-improved policy, we apply simulation. Instead of computing the actions $\widetilde{\pi}^a$ and $\widetilde{\pi}^b$ for all possible state (Which may be prohibitively many), we only compute them for states visited in the simulation.

5.4 Results

We compare the performance of the different fulfilment strategies for a wide range of instances. We first introduce a base test case, and compare the different fulfilment strategies on economic performance and fill rates. We then investigate the different fulfilment settings by varying the parameters of the base test case. The data set used is based on recent literature studying similar omni-channel retailer network settings (Xu & Cao, 2019; Bayram & Cesaret, 2021; Goedhart et al., 2023).

5.4.1 Data and design of experiments

For the base test case we assume that the stores differ in average walk-in demand and average BOPS demand: there are small stores with walk-in demand $\mu_{1,j} = 2$ and BOPS demand $\mu_{2,j} = 1$, medium stores with walk-in demand $\mu_{1,j} = 4$ and BOPS demand

$\mu_{2,j} = 2$, and large stores with walk-in demand $\mu_{1,j} = 6$ and BOPS demand $\mu_{2,j} = 3$. Furthermore, the average BODH demand is $\mu_0 = 60$. The walk-in, BOPS, and BODH demand is assumed to be independently Poisson distributed, and as we need a finite support they are right-truncated at a cumulative probability of 99.9% as described in Cohen (1954).

Furthermore, we assume there are three different types of stores which differ on the number of BOPS and walk-in demand they receive. For the base test case we set the number of small stores as $m_1 = 10$, the number of medium stores as $m_2 = 10$, and the number of large stores as $m_3 = 10$. We denote the number of different types of stores with $M = \{m_1, m_2, m_3\}$.

The stores have varying walk-in and BOPS demand but all have identical review periods, lead time, price, and costs, which is often the case in practice (e.g., Noordhoek et al., 2018). We assume days as sub-periods and a review period of 7 days ($R = 7$) to reflect a weekly ordering decision, with a fixed lead time of $L = 2$, indicating that for example if orders are placed every Monday $(t = 1)$ these are delivered on Wednesday $(t = 3)$.

The price of the product is fixed at $p = 100$, the procurement cost c is 70. The cost of shipping an BOPS or BODH order from the OFC or store is set to $k = 5$. We assume that the picking and packing of an BOPS or BODH order at the store is more expensive than at the OFC, as OFC are designed to efficiently handle orders. Thus the handling cost at store $j > 0$ is $u_j = 5$ while at the OFC $u_0 = 2.5$. Furthermore, store shelves are more expensive than those of OFC, therefore the holding cost at store $j > 0$ is $h_j = 1$ and at the OFC $h_0 = 0.5$. The penalty cost of cancelling an BOPS or BODH order is $g = 10$.

For further analyses, twenty instances are created by varying subsets of parameters from the base test case, as shown in Table 5.2. For the instances in which the number of stores varies, we ensure that the total demand in each channel remains equal.

In the remainder of this paper, if an instance does not specify a certain parameter value, it will be equal to its base test case value. The results were obtained by implementing the model in Python version 3.7.2, which was run on a Personal Computer with Intel Xeon W-2133 CPU @ 3.60 GHz and 32GB of RAM. The economic performance and service levels were obtained by simulating the policies for 100,000 periods with a small warm-up period.

Table 5.2: Parameter values, where the base case is given in bold.

Parameter	Symbol	Values
In-store demand	$(\mu_{1,j}, \mu_{1,j}, \mu_{1,j})$	$\mathbf{(2, 4, 6)}$
BOPS demand	$(\mu_{2,j}, \mu_{2,j}, \mu_{2,j})$	$\mathbf{(1, 2, 3)}$
BODH demand	μ_0	$\mathbf{60}$
Review period	R	$3, \mathbf{7}$
Lead time	L	$\mathbf{2}$
Number of stores	$\{m_1, m_2, m_3\}$	$(\mathbf{\{10, 10, 10\}}, \{0, 30, 0\}, \{60, 0, 0\}, \{0, 0, 20\}, \{19, 10, 7\}, \{13, 10, 9\}, \{7, 10, 11\}, \{1, 10, 13\})$
Price	p	$\mathbf{100}$
Procurement cost	c	$\mathbf{70}$
Shipping cost	k	$2.5, \mathbf{5}, 7.5$
Penalty cost	g	$\mathbf{10}$
Picking and packing cost	u_j	$0, \mathbf{2.5}, \mathbf{5}, 7.5$
Holding cost	h_j	$\mathbf{0.5}, \mathbf{1}, 1.5$

5.4.2 Analysis of the base test case

Economic evaluation

First, we evaluate the base test case on different economic indicators for the four different fulfilment options, based on profits as well as underlying revenues and costs. We define the performance as the percentage deviation of the economic performance of the fulfilment option of LF, OC, and FF in relation to SF. For instance, a positive percentage of 4.4 for the profit of FF indicates that this fulfilment strategy has a 4.4% higher profit than SF. Table 5.3 presents the economic performance for the different fulfilment strategies.

From Table 5.3 it is observed that FF has the best performance in terms of profits, which is expected as it has the most options to fulfil demand. SF has the lowest profits: it has relatively high revenues and the benefit of pooling all demand at the stores, but the store holding, and picking and packing cost outweigh the cost savings of demand pooling. OC outperforms LF, indicating that the benefits of pooling BOPS demand at the OFC outweigh the extra cost made of fulfilling the demand via the OFC.

SF has the highest revenue from walk-in demand, which is expected as the stores have high stock levels as they need to fulfil multiple types of demand. LF has slightly lower walk-in revenue as the stock levels at the store are lower since it is not used for the BODH demand. OC show the lowest revenue from walk-in demand, which is expected

Table 5.3: Results of the economic performance for the base test case for store fulfilment (SF), local fulfilment (LF), online centralised fulfilment (OC), and flexible fulfilment (FF).

			Performance (%)		
	Configuration	SF	LF	OC	FF
	Profit	35717.51	1.61	2.17	4.14
Revenue	Walk-in	81664.05	-1.87	-2.82	-0.38
	BOPS	40425.56	-3.75	0.37	0.38
	BODH	41819.04	-1.00	-0.47	0.02
Costs	Replenishment	114736.06	-2.11	-1.44	-0.09
	Holding	7235.90	-9.01	-20.00	-15.36
	BOPS fulfilment	2021.28	-3.75	50.56	28.72
	BODH fulfilment	4181.90	-25.75	-25.35	-23.36
	Penalty	0.83	5006.02	2360.30	-97.09

as the store stock levels in this setting are the lowest. In FF, the retailer can allocate orders to the store or OFC, and is therefore able to protect store inventory from BODH demand. Thus the walk-in revenue is high, but not as high as SF which has higher store stock levels.

For the revenue from BOPS demand, LF has the lowest revenue as it can only fulfil the BOPS demand from the store stock. Although SF also has limited options for fulfilling the BOPS demand it has the advantage of higher stock levels at the stores. SF and FF have the highest BODH revenue, which is expected as they have more flexibility in fulfilling the BODH demand than in design LF and OC, namely the stores and OFC. LF has the lowest revenue from the BODH demand, as BODH demand is not pooled with another type of demand resulting in less safety stock at the OFC.

The total replenishment costs are the highest for SF, the BODH demand is distributed over all stores therefore each location has to hold extra inventory and safety stock. In LF the BODH demand is pooled at the OFC resulting in a lower safety stock than at SF. In OC there is most demand pooling at one location thus a further decrease in safety stock and replenishment. FF has similar to SF in that demand is distributed over more locations thus lower demand pooling, resulting in more safety stock thus higher replenishment decision. Additionally, as FF allows for more fulfilment options excessive stock can more easily be handled, thus replenishing more does not result in overstocking.

When the stores are more involved in the fulfilment of BOPS and BODH demand, their stock positions are higher to satisfy those demands. This also results in higher

Table 5.4: Fill rate on the day before replenishment for the base test case for store fulfilment (SF), local fulfilment (LF), online centralised fulfilment (OC), and flexible fulfilment (FF).

Configuration	SF	LF	OC	FF
Walk-in	0.998	0.883	0.839	0.974
BOPS	0.974	0.759	1.000	1.000
BODH	0.999	0.929	0.966	1.000
Total	0.992	0.864	0.911	0.987

holding costs, as the fulfilment strategies in which stores are more involved in the BODH channel (e.g., SF and LF) have the highest holding costs.

The BOPS fulfilment cost is the lowest for SF and LF, as there the store is used for the fulfilment which has the lowest cost. For LF the cost is a bit lower, due to fulfilling fewer BOPS demand as it has lower stock levels. OC has the highest BOPS fulfilment costs as all demand is fulfilled from the OFC which has high fulfilment cost. FF prefers to use store inventory for the BOPS fulfilment therefore it is lower than compared to OC, where BOPS demand is fulfilled via the OFC.

LF and OC have similar BODH fulfilment costs, as both only use the OFC for the BODH demand. FF sometimes uses the stores for the fulfilment, which is costlier, similarly SF only uses stores and thus has the highest BODH fulfilment costs.

LF and OC have the highest penalty costs (for not being able to fulfil BOPS or BODH orders), which is expected as they have lower inventory positions. Therefore, the retailer will encounter stock-outs and has to cancel orders. FF has almost no penalty cost, indicating that there is always a product available for the fulfilment of BOPS or BODH demand in the network. This is expected as in FF, both stores and the OFC can fulfil any BOPS or BODH demand.

Fill rate

Besides economic performance, an important performance indicator is the fill rate, which is the fraction of demand that is being fulfilled with the available stock. We are mostly interested in the fill rate on the day before replenishment, as then stock levels are low and the effects of improper inventory management is most notable. For each fulfilment strategy, we report the fill rate before replenishment for each type of demand in Table 5.4.

From Table 5.4, it is observed that SF has the highest fill rate due to the high stock levels at stores. The fill rate of the BODH channel is higher than the BOPS channel as BODH demand can be fulfilled from any store in the network as opposed to BOPS demand. Thus only when there is a network-wide stock-out BODH demand cannot be fulfilled.

LF has the lowest BOPS fill rate, as the store is only used for walk-in and BOPS demand and thus has lower inventory levels. Since walk-in demand is fulfilled before BOPS demand the retailer might not have enough stock to fulfil these. Similarly, the BODH fill rate is lower as the OFC only holds stock for BODH demand, thus has less safety stock resulting in more frequent stock-outs.

OC fulfils all BOPS demand, which is a result of the order in which BOPS and BODH demands are fulfilled: BOPS demand is fulfilled before BODH demand is satisfied. To prevent BODH demand to be cancelled and an expensive penalty cost being incurred, the OFC has a high enough stock level. The walk-in fill rate is the lowest among all fulfilment strategies as the store has low stock levels due to only serving walk-in demand.

FF fulfils all BOPS and BODH demands due to the ability to fulfil these demands from many different locations. Only with a network-wide stock-out the fill rate would be below one, however by appropriately fulfilling the demand with the different locations throughout the week this can be avoided. The fill rate of the walk-in demand is high due to the demand pooling effect that occurs at the store partly, it is not as high as in SF as not all demand is pooled at the store.

The results of the base test case show that SF has the lowest profitability, although it allows for demand pooling at the store level and thus achieves a high fill rate at the expense of high costs. LF has less demand pooling as the BODH demand is fulfilled via an OFC, thus the fill rate is much lower. However, as the OFC is less costly for fulfilling BODH demand it achieves higher profit. OC has a high performance, as BOPS and BODH demands are pooled at the OFC which has low holding cost. However as there is no pooling at the store level, the fill rate for walk-in demand is the lowest. FF has all the advantages described above without being influenced much by the negative consequences. It has the highest profit and can fulfil all BOPS and BODH demands.

5.4.3 Analysis of the full set of instances

The results of the different fulfilment strategies are evaluated for a wide range of instances as described in Section 5.4.1. We evaluate them on their profit and total fill rate before replenishment. The first performance measure gives us an indication of which fulfilment strategy performs best under different network configurations and cost settings. The second performance measure indicates the performance of the demand fulfilment. Furthermore, we evaluate for different parameters which fulfilment strategies have the highest performance and evaluate where trade-offs between networks occur.

Economic evaluation and fill rate

Table 5.5 presents the economic performance and fill rate before replenishment for the different instances. We present the economic performance as the average profit made in a review period for SF, for the other networks we again present the performance as the percentage deviation from SF. A positive percentage thus indicates a higher profit compared to SF.

It is observed from Table 5.5 that (as we saw for the base test case) FF has the highest profits. This is expected as it can use the same fulfilment policies of all three other networks, thus should always perform equally or better than the others. Similar to the base test case, SF has almost always the highest fill rate except for the instance with a high procurement cost. Also, we see that across the instances, LF has the lowest fill rates, and FF has fill rates similar or just below SF.

The instances with different numbers of stores are set to ensure that the average demand for the different fulfilment options remains the same. We observe that with fewer stores in the network, all fulfilment strategies have higher average profits and the performance between the fulfilment strategies increases. With fewer stores, demand is pooled over fewer locations, thus the demand pooling effect is higher resulting in better performance.

With decreasing holding costs for stores, fulfilment strategies that focus on storing products at OFC are negatively influenced. Therefore OC performs worst with lower store holding cost, but vice versa if the holding cost at stores are high it performs better. As for FF, it can easily store products at the least costly locations and thus can take advantage of lower holding cost to improve its performance significantly. Higher store

Table 5.5: Economic evaluation and fill rate before replenishment of all instances for store fulfilment (SF), local fulfilment (LF), online centralised fulfilment (OC), and flexible fulfilment (FF).

	Profit				Fill rate			
		Performance (%)						
	SF	LF	OC	FF	SF	LF	OC	FF
Base test case*	35717.51	1.61	2.17	4.14	0.99	0.86	0.91	0.99
$M = \{0, 30, 0\}$	35592.67	1.61	2.21	4.23	0.99	0.86	0.91	0.99
$M = \{60, 0, 0\}$	34900.82	0.40	1.37	3.66	0.98	0.84	0.90	0.97
$M = \{0, 0, 20\}$	36039.62	2.08	2.46	4.42	1.00	0.88	0.91	0.99
$M = \{19, 10, 7\}$	35636.82	1.10	1.75	3.76	0.99	0.86	0.91	0.98
$M = \{13, 10, 9\}$	35629.37	1.61	2.20	4.19	0.99	0.86	0.91	0.99
$M = \{7, 10, 11\}$	35824.07	1.55	2.08	4.04	0.99	0.87	0.91	0.99
$M = \{1, 10, 13\}$	35883.82	1.87	2.34	4.29	0.99	0.87	0.91	0.99
$h_{j>0} = 0.5$	38928.22	1.21	-0.76	1.49	1.00	0.93	0.96	1.00
$h_{j>0} = 1.5$	32507.18	2.92	6.37	8.02	0.91	0.79	0.87	0.96
$k = 2.5$	36693.66	1.73	5.05	5.38	0.99	0.86	0.91	0.99
$k = 7.5$	34733.28	1.51	-0.85	3.11	0.99	0.86	0.91	0.98
$u_{j>0} = 2.5$	37671.17	-1.04	-3.13	0.52	1.00	0.87	0.91	0.98
$u_{j>0} = 7.5$	33747.67	4.65	8.13	8.34	0.99	0.86	0.91	0.99
$u_0 = 0$	35717.51	4.51	7.92	8.07	0.99	0.86	0.91	0.99
$u_0 = 5$	35717.51	-1.29	-3.58	0.28	0.99	0.86	0.91	0.98
$c = 55$	35110.29	3.32	3.81	5.22	1.00	0.91	0.94	1.00
$c = 65$	35468.11	2.37	2.87	4.58	1.00	0.88	0.92	0.99
$c = 75$	35924.52	0.84	1.52	3.86	0.97	0.83	0.89	0.98
$c = 85$	33860.59	4.67	6.68	9.01	0.56	0.68	0.80	0.84
$R = 3$	16474.83	-0.04	-0.10	0.04	1.00	0.70	0.82	0.97
Average	34656.15	1.77	2.41	4.37	0.97	0.85	0.90	0.98

*$M = \{10, 10, 10\}, h_{j>0} = 1, k = 5, u_{j>0} = 5, u_0 = 2.5, c = 70, R = 7$

holding cost negatively influence the fill rate, as products become more expensive to hold the safety stock is decreased resulting in more frequent stock-outs.

OC is mostly influenced by different shipping costs, as the BOPS demand fulfilment in this setting is shipped from the OFC and cannot be fulfilled by the store. With low shipping cost OC has almost the same performance as FF, indicating that it is more profitable to fulfil BOPS demand from the OFC than stores. Although higher shipping cost result in BODH demand being less profitable, the total fill rate remains roughly equal. The reason for this is that the retailer tries to avoid incurring a penalty cost of cancelling an BODH demand, thus although the profit margin is relatively low it still tries to fulfil all BOPS and BODH demands.

With lower in-store picking cost, the fulfilment strategies that rely on the stores for the fulfilment of BOPS and BODH demands have a higher profit. An opposite trend is observed for higher picking cost. With high in-store picking cost OC has almost the same performance as FF, as the stores are becoming too costly to fulfil BOPS demand the OFC is preferred.

For the OFC picking cost, a similar trend is observed, having low OFC picking cost makes the OFC more profitable to be utilised for BOPS and BODH demands and vice versa for high picking cost. Therefore at a low OFC picking cost OC has almost similar performance to FF. However, at high OFC picking cost OC performs worst due to only being able to fulfil BOPS and BODH demands via the OFC. LF also is negatively influenced by high OFC picking cost and performs worse than when only stores are used for fulfilment.

With lower procurement cost, the profit margins per product sold increases. This also results in higher fill rates as extra safety stocks become more profitable. Higher procurement cost negatively influence SF the most, as the profit margins are already low at the store due to higher holding costs.

With a shorter review period, LF and OC have a lower profit than SF. In these cases, the overall inventory levels are lower as the retailer orders more often. There thus the benefit of low holding cost at the OFC is decreased as products are stored for shorter. This causes the advantage of pooling the demand at the stores to outweigh the benefit of lower cost at the OFC, resulting in SF outperforming LF and OC.

Trade-off between fulfilment strategies

It is clear that FF always outperforms the other fulfilment options on profit, because FF can use the same fulfilment policies as the other strategies. However, it is observed that for the other fulfilment strategies that different holding and picking costs influence which fulfilment strategy is the most profitable. Therefore, for different costs we identify which fulfilment strategy has the highest performance. As we are only interested in SF, LF, and OC (since FF will always outperform the others), we only look at the performance of these settings. Figure 5.4 shows which fulfilment strategy has the highest performance for different picking costs for the store and OFC (a) and different holding costs for the store and OFC (b). The trade-off lines were found via a grid search over all the possible combinations of cost.

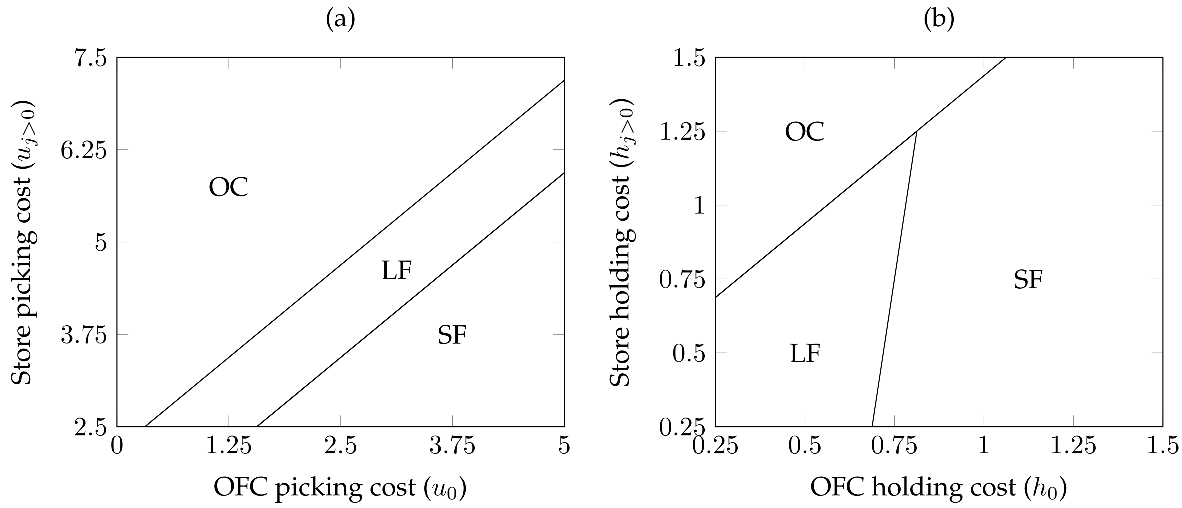


Figure 5.4: Profit trade-off of the fulfilment strategies for different picking cost (a) or holding cost (b) for store fulfilment (SF), local fulfilment (LF), and online centralised fulfilment (OC).

From Figure 5.4(a) it is clear that when the cost of picking orders from store stock is low, SF has the highest performance. Low picking cost at the stores make store fulfilment more attractive. When the ratio between OFC and store picking costs decreases, store fulfilment becomes less profitable and using the OFC for the fulfilment of demand becomes more attractive. Therefore, LF becomes more profitable, as fulfilling BODH demand from the OFC and BOPS demand from the stores has the lowest picking cost. When the store picking cost becomes too high compared to the OFC picking cost, fulfilling BOPS demand from the OFC becomes more profitable. Therefore then OC has the highest performance, as the BOPS and BODH demands are then fulfilled from the OFC.

Figure 5.4(b) shows the trade-off between different fulfilment strategies for different holding costs of the OFC and stores. A similar trend for the picking costs is observed, where if the store holding cost is low and the OFC holding cost high, SF has the highest performance. For high store holding cost and low OFC holding cost OC has the highest performance, as then most products are stored at the OFC. When the store and the OFC have similar holding costs the different handling and picking costs between stores and OFC decide which fulfilment strategy has the highest performance. For LF the fulfilment of BOPS and BODH demands is the lowest, therefore has the highest performance. However, if the holding cost of the OFC exceeds a certain threshold LF does not have the highest performance and is outperformed by SF. At this threshold the advantage of pooling BODH demand at the stores outweighs the cost savings of

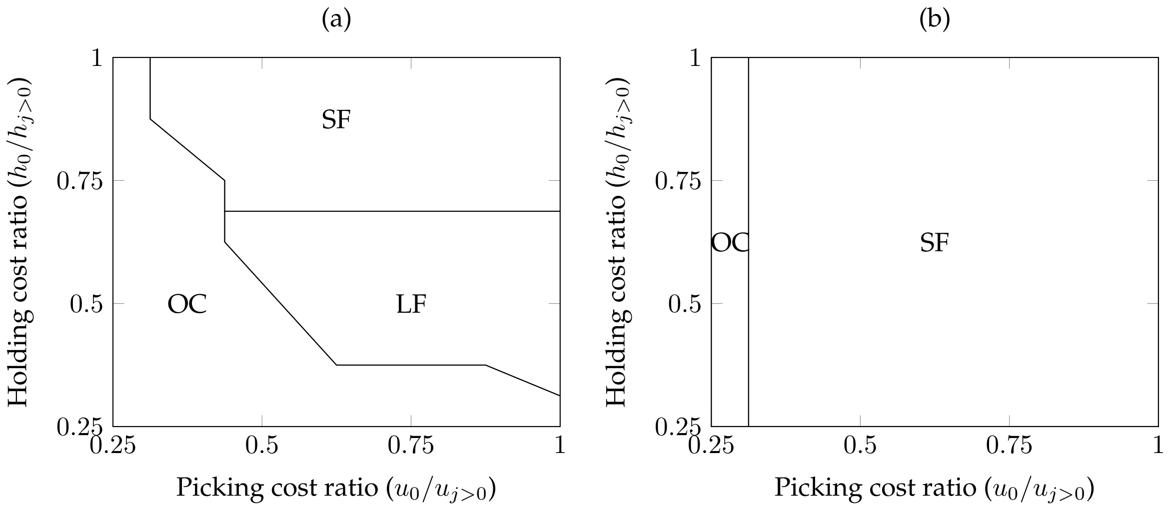


Figure 5.5: Profit (a) and fill rate before replenishment (b) trade-off of the fulfilment strategies for different ratios of picking and holding cost for store fulfilment (SF), local fulfilment (LF), and online centralised fulfilment (OC).

storing products at the OFC, resulting in SF having a higher performance.

When investigating the fill rate before replenishment in Figure 5.4(a) and (b), it was observed that similar to the results found in Table 5.5 SF has the highest performance for all different cost combinations.

To study the interactions between picking and holding costs, we can use the ratio between the OFC and store cost parameters for each. Figure 5.5 shows which fulfilment strategy has the highest profit (a) and fill rate before replenishment (b) for different holding and picking costs. The ratio between the holding costs of the OFC and store ($h_0/h_{j>0}$) is represented on the horizontal axis and the ratio between the picking costs of the OFC and store ($u_0/u_{j>0}$) is represented on the vertical axis. We find different ratios by varying the cost of the stores and keeping the costs of the OFC constant.

From Figure 5.5(a), we see that SF performs best when the cost difference between the OFC and store for both types of cost is low. As store picking cost increases, fulfilment via the stores becomes less profitable, and OC has better performance (because there are no orders picked at the store).

When the holding cost of the store or the cost of picking demand at the store becomes higher, OC is the most profitable. With high store operating costs using the OFC for the fulfilment of demand becomes more interesting. With lower store holding cost and decreasing store picking cost the fulfilment strategy with the highest performance shifts from OC to LF to SF. This is the result of the store picking cost making the store more profitable to be used for the fulfilment of demand. The in-between region where LF is most profitable is the result of BOPS demand preferred to be fulfilled by the store due to lower shipping cost, however the low OFC holding cost make BODH demand to be fulfilled from the OFC less costly.

Figure 5.5(b) shows that SF has the highest fill rate in most cases, which is in line with the findings of Table 5.5 as the stores have stock levels as they have to fulfil multiple types of demand. However, when the cost of storing a product at the store level are relatively high, OC has higher fill rates. With higher holding cost at the store, the safety stock becomes more expensive and thus will be decreased. This results that SF has lower fill rates since OC mostly uses the OFC for holding products.

The performance trade-off between the fulfilment strategies is dependent on the cost structure of the studied problem. When store costs are relatively low or OFC costs relatively high, we see that SF has the highest profit compared to LF and OC. With relatively high store costs or low OFC costs, OC has the highest profit compared to SF and LF. The overall fill rate for SF is the highest for almost all fulfilment options due to the demand pooling effect at each store, as each store has to fulfil three different types of demand.

5.5 Conclusion and Discussion

With the growth of omni-channel retailing, many retailers have to rethink their supply chain network, due to the different ways customers want to receive their goods. For online orders, the retailer can decide how to fulfil the demand, either from store inventory or from a dedicated OFC. As different fulfilment locations have different fulfilment-related costs, the choice for an omni-channel fulfilment strategy is not obvious. This paper contributes to the literature by investigating different fulfilment strategies and their underlying replenishment and demand fulfilment decisions for multiple stores. We formulate the model as a Markov Decision Process and solve this via a one-step policy improvement approach.

Our results confirm that the most flexible network setup, i.e. allowing both the stores and the OFC to fulfil BOPS and BODH demands, leads to the highest total profit. However, when the cost of fulfilling via stores becomes more than double the cost than via the OFC, then only using the OFC for fulfilling BOPS and BODH demands has almost the same performance as using both the stores and the OFC. When the fulfilment cost via stores is equal to the fulfilment cost of an OFC, then the added value of an OFC decreases. In addition, when only store inventory is used for the fulfilment of demand, our results show an expected lower profit, but higher fill rates than in situations in which an OFC is also used. Furthermore, only using the OFC for BOPS and BODH demand can result in lower profits and lower in-store fill rates. In general, the results confirm that only having one location available for the fulfilment of BOPS and BODH orders limits demand fulfilment flexibility. The extent in which this impacts profitability and service levels does however depend on the relative cost differences between inventory and fulfilment costs at the store level and the OFC level. The results show that when the cost of the store increases and/or the cost of the OFC decreases, store fulfilment becomes less profitable while using the OFC is more profitable. When demand can be fulfilled more flexibly or is pooled at a single location, the fill rates increase. However, this might be at the expense of higher costs.

From the results, several managerial insights can be derived. When stores are used for the fulfilment of online demand, the inventory positions of the stores need to be increased to fulfil these extra demands. The advantage of this is that this results in higher fill rates for in-store customers at the expense of higher store holding costs. However, when the cost of fulfilling via a store becomes almost double that of fulfilment via the OFC (due to e.g. more expensive store personnel or higher inventory costs), then having stores act as fulfilment centres has little added value. Contrary, when the costs of stores and OFC are identical than the added value of an OFC is negligible.

Although our omni-channel retail network model captures most typical characteristics of omni-channel retailers, some aspects were not included in our research. First, our model does not include transshipment between stores. A retailer can decide to tranship products proactively between stores to reduce stock-outs, which can result in decreased inventory imbalance across the retailer's network and an increase of in-store revenue. Secondly, in order to reduce the complexity of the model we did not account for the fact that customers might switch between channels when encountering a stock-out in store. On average, our results show high store fill rates, meaning that including this behaviour might not impact the results much. However, in some settings, this kind of customer behaviour might be relevant and change the decision-making of the retailer. Third, possible product returns can further complicate inventory management.

Products that are ordered online might be returned to stores, resulting in unwanted store inventory imbalances. Considering cross-channel sales returns is therefore an interesting direction for further research. Fourth and finally, we assumed that demand is stationary and non-seasonal, which holds for several but surely not all products. Investigating the studied problem for non-stationary seasonal products might be another interesting further research direction, as often a clearance sale is done at the end of a season where the profit margins becomes smaller. This might induce the retailer to pro-actively reduce inventory before this period to avoid losing profit margins.

Appendix 5.A Different fulfilment strategies

Store Fulfilment (SF)

In the setting where only stores are used to fulfil BOPS and BODH orders, the OFC does not exists. To adapt the described model above to reflect having no OFC we set the ordering decision of the OFC to zero, thus $Q_0 = 0$. This ensures there are no products available at the OFC, thus it cannot fulfil any orders.

Local Fulfilment (LF)

In the setting where demand is fulfilled locally, the store has to fulfil the BOPS demand and the OFC only fulfils the BODH demand. As the OFC only fulfils BODH demand the inventory state space and ordering action space is reduced to $\mathcal{I}_0 = \{0, 1, \ldots (R+L) \cdot D_0\}$ and $Q_0 \in \mathcal{Q}_0(t, I_0)$ where:

$$\mathcal{Q}_0(t, I_0) = \begin{cases} \{0, 1, \ldots, (R+L) \cdot D_0 - I_0\} & \text{if } t < L \\ 0 & \text{otherwise.} \end{cases}$$

As the store does not need to fulfil any BODH demand the state space and ordering action space of the store is reduced to $\forall j \in \{1, 2, \ldots, J\} : \mathcal{I}_j = \left\{0, 1, \ldots, (R+L) \cdot \left(\sum_{i=1}^{2} D_{i,j}\right)\right\}$ and $Q_j \in \mathcal{Q}_j(t, I_j)$ where:

$$\mathcal{Q}_j(t, I_j) = \begin{cases} \left\{0, 1, \ldots, (R+L) \cdot \left(\sum_{i=1}^{2} D_{i,j}\right) - I_j\right\} & \text{if } t < L \\ 0 & \text{otherwise.} \end{cases} \quad \forall j \in \{1, 2, \ldots, J\}$$

Furthermore, stores do not fulfil any BODH orders, thus $\forall j \in \{1, 2, \ldots, J\} : a_j = 0$.

As the OFC does not fulfil any BOPS orders the action space of the BOPS fulfilment is adapted to $b_0 = 0$.

Online Center Fulfilment (OC)

In the setting where the OFC fulfils all BOPS and BODH orders the inventory state space and ordering action space of the store is reduced as it only needs to fulfil in-store demand. Thus $\forall j \in \{1,2,\ldots,J\} : \mathcal{I}_j = \{0,1,\ldots,(R+L)\cdot D_{1,j}\}$ and

$$\mathcal{Q}_j(t, I_j) = \begin{cases} \{0,1,\ldots,(R+L)\cdot D_{1,j} - I_j\} & \text{if } t < L \\ 0 & \text{otherwise.} \end{cases} \quad \forall j \in \{1,2,\ldots,J\}$$

As the stores do not fulfil any BOPS or BODH order the action for these decisions is set to zero, thus $\forall j \in \{1,2,\ldots,J\} : a_j = 0$ and $\forall j \in \{1,2,\ldots,J\} : b_j = 0$.

Appendix 5.B Expected immediate rewards

Expected immediate rewards of epoch 1

In the first epoch the expected immediate rewards consists of the revenue from products sold in-store, the holding cost, and ordering cost. The expected immediate rewards in epoch 1 given the current state and ordering action is calculated as follow:

$$C_1(s,q) = p \sum_{j=1}^{J} \left(\sum_{d_{1,j} < I_j} d_{1,j} \cdot P_{1,j}\left(d_{1,j}\right) + \sum_{d_{1,j} \geq I_j} I_j \cdot P_{1,j}\left(d_{1,j}\right) \right) - \sum_{j \in \mathcal{J}} \left(h_j \cdot I_j - c \cdot \delta(t=1) \cdot q_j\right), \tag{5.6}$$

the first term reflects the sales from all consumers who purchased their product in-store, where p is the price of the product. The second term is the the holding costs (h_j) of the OFC and store. The last term reflects the procurement cost of a product (c), which is only incurred at the beginning of the period.

Expected immediate rewards of epoch 2

The immediate rewards in epoch 2 consists of the revenue from fulfilling BOPS orders and a penalty cost for every BOPS order that is not fulfilled. Furthermore, there is handling cost for preparing the BOPS order in-store, and when the BOPS orders is shipped from the OFC a shipping and handling cost from the OFC to the store is incurred:

$$C_2\left(s', b\right) = \sum_{j=1}^{J} \left(p - u_j\right) \cdot b_j + \left(p - \left(u_0 + k\right)\right) \cdot b_0 - g \cdot \left(\sum_{j=1}^{J} \left(B_j - b_j\right) - b_0 \right) \tag{5.7}$$

Where the first term is the profit of fulfilling a BOPS order from store inventory where a cost is incurred for the in-store handling (u_j) for preparing the BOPS order. The second term is the profit from fulfilling a BOPS order from the OFC inventory where a cost is incurred for the handling and shipping ($u_0 + k$) for preparing the BOPS order. The last term reflects the penalty cost (g) incurred for not satisfying a BOPS orders.

Expected immediate rewards of epoch 3

The immediate rewards in epoch 3 consists of the revenue from fulfilling online orders minus the fulfilment cost of the store and OFC, and a penalty cost for every online order that is not fulfilled:

$$C_3\left(s'', a\right) = \sum_{j \in \mathcal{J}} \left(p - \left(u_j + k\right)\right) \cdot a_j - g\left(A - \sum_{j \in \mathcal{J}} a_j\right) \tag{5.8}$$

Where the first term is the profit of fulfilling an online order, where a cost is incurred for the handling and shipping cost ($u_j + k$). The second term reflects the penalty cost incurred for not satisfying an online order.

Appendix 5.C State space

The state space of the period is limited by the review period: $t \in \{0, 1, \ldots, R-1\}$.The state space of the epoch consists of the beginning, middle and end of the sub-period: $e \in \{1, 3, 2\}$.

The state space of the inventory level of the OFC is limited by the maximum expected orders it needs to fulfil during the review period plus lead time. The state space of the inventory of the OFC is defined as $I_0 \in \mathcal{I}_0$ with $\mathcal{I}_0 = \left\{0, 1, \ldots, (R+L) \cdot \left(\sum_{j=1}^{J} D_{2,j} + D_0\right)\right\}$. We take the maximum expected BOPS demand of each store as the OFC can fully fulfil these.

The state space of the inventory levels of each store in the network is limited by the maximum expected demand of the store during the review period plus lead time. The state space of the inventory level of a store is defined as $I_j \in \mathcal{I}_j$ with $\forall j \in \{1, 2, \ldots, J\}$: $\mathcal{I}_j = \left\{0, 1, \ldots, (R+L) \cdot \left(\sum_{i=1}^{2} D_{i,j} + D_0\right)\right\}$. Reason we take the maximum expected BODH demand is that all BODH orders can be fully fulfilled from the store.

The state space of the outstanding orders for the OFC is dependent on the current sub-period, inventory position, and the expected demand during review period plus lead time and defined as $Q_0 \in \mathcal{Q}_0(t, I_0)$ where:

$$\mathcal{Q}_0(t, I_0) = \begin{cases} \left\{0, 1, \ldots, (R+L) \cdot \left(\sum_{j=1}^{J} D_{2,j} + D_0\right) - I_0\right\} & \text{if } t < L \\ 0 & \text{otherwise.} \end{cases}$$

The state space of the outstanding orders for the stores is dependent on the current sub-period, inventory position, and the expected demand during review period plus lead time and defined as $Q_j \in \mathcal{Q}_j(I_j)$ where:

$$\mathcal{Q}_j(t, I_j) = \begin{cases} \left\{0, 1, \ldots, (R+L) \cdot \left(\sum_{i=1}^{2} D_{i,j} + D_0\right) - I_j\right\} & \text{if } t < L \\ 0 & \text{otherwise.} \end{cases} \quad \forall j \in \{1, 2, \ldots, J\}$$

The state space for the accepted BOPS demand is limited by the total BOPS expected BOPS demand on a day. It is defined as $B_j \in \mathcal{D}_{2,j}$ with $\mathcal{D}_{2,j} = \{0, 1, \ldots, D_{2,j}\}$.

The state space for the accepted BODH demand is limited by the total expected BODH demand on a day and defined as $A \in \mathcal{D}_0$ with $\mathcal{D}_0 = \{0, 1, \ldots, D_0\}$.

Appendix 5.D Action space

Action space of epoch 1

At the beginning of the period a replenishment decision for the OFC and each store is made q_j. The action space of the replenishment decision is equal to that of the ordering state $q_j \in \mathcal{Q}_j\left(t, I_j\right)$. The total action and action space of the replenishment decision is given as: $\mathcal{Q}\left(t, I\right) = \prod\limits_{j \in \mathcal{J}} \mathcal{Q}_j\left(t, I_j,\right)$, thus the vector consisting of all the replenishment decisions has an actions space of $q \in \mathcal{Q}\left(t, I\right)$.

Action space of epoch 2

At the beginning of epoch 2 the retailer has to allocate the accepted BOPS orders to each fulfilment location. b_j is the amount of BOPS orders that are fulfilled from either the OFC $j = 0$ or the store $j > 0$. For the OFC the BOPS fulfilment decision is limited by the closing inventory of the OFC and the total number of BOPS orders, thus $b_0 \in \mathcal{B}_0\left(I_0', B'\right)$ with $\mathcal{B}_0\left(I_0', B'\right) = \left\{0, 1, \ldots, \min\left(I_0', \sum_{j=1}^{J} B_j'\right)\right\}$. For the store the BOPS fulfilment decision is limited by the closing inventory of the store and the number of BOPS orders the store received, thus $b_j \in \mathcal{B}_j\left(I_j', B_j'\right)$ with $\forall j \in \{1, 2, \ldots, J\}$: $\mathcal{B}_j\left(I_j', B_j'\right) = \left\{0, 1, \ldots, \min\left(I_j', B_j'\right)\right\}$. Furthermore b_0 and b_j are restricted by the following constraint: $\sum_{j=1}^{J}(B_j - b_j) \geq b_0$ as no more BOPS orders can be fulfilled from the OFC than there are available. The total action and action space of the BOPS fulfilment decision is given as: $\mathcal{B}\left(I', B'\right) = \prod_{j \in \mathcal{J}} \mathcal{B}_j\left(I_j', B_j'\right)$, thus the vector consisting of all the BOPS fulfilment decisions has an actions space of $b \in \mathcal{B}\left(I', B'\right)$.

Action space of epoch 3

At the beginning of epoch 3 the retailer has to allocate the collected online orders of the day to each store and the OFC. As any store or OFC can fulfil any online order the allocation decision has to be made for each store or OFC, we denote the decision as a_j. The allocation decision is limited by the closing inventory of the OFC or store, and the online demand thus $a_j \in \mathcal{A}_j\left(I_j'', A''\right)$ with $\mathcal{A}\left(I_j'', A''\right) = \left\{0, 1, \ldots, \min\left(I_j'', A''\right)\right\}$. Furthermore a_j is restricted by the following constraint: $\sum_{j \in \mathcal{J}} a_j \leq A''$ as no more online orders can be fulfilled than there are available. The online order allocation action space of all stores combined is given by $\mathcal{A}\left(I'', A''\right) = \prod_{j \in \mathcal{J}} \mathcal{A}_j\left(I_j'', A''\right)$, thus the vector consisting of all allocation decisions has an actions space of $a \in \mathcal{A}\left(I'', A''\right)$.

Appendix 5.E Value iteration for the stores.

Algorithm 6: Value iteration for the stores.

initialisation: $v_{0,j} = 0$;

for $j \in \{1, 2, \ldots, J\}$ **do**

  for $n = 1 : 1 : n_{max}$ **do**

    for $t = 7 : -1 : 1$ **do**

      for $s''_j = \left(t, e = 3, I''_j, Q''_j, B''_j = 0, A''\right) \in \left\{\left(t, e = 3, I''_j, Q''_j, B''_j = 0, A''\right) \mid I''_j \in \mathcal{I}_j, Q''_j \in \mathcal{Q}_j\left(t, I''_j\right), A'' \in \mathcal{D}_0\right\}$ **do**

$$v_{n,j}\left(s''_j\right) = \max_{a_j \in \mathcal{A}\left(I''_j, A''\right)} \left\{ C_3\left(s''_j, a_j\right) + v_{n-1,j}\left(s'''_j\right) \right\}$$

        where: s'''_j is from equation (5.15) and (5.16)

      end

      for $s'_j = \left(t, e = 2, I'_j, Q'_j, B'_j, A'\right) \in \left\{\left(t, e = 2, I'_j, Q'_j, B'_j, A'\right) \mid I'_j \in \mathcal{I}_j, Q'_j \in \mathcal{Q}_j\left(t, I'_j\right), B'_j \in \mathcal{D}_{2,j}, A' \in \mathcal{D}_0\right\}$ **do**

$$v_{n,j}\left(s'_j\right) = \max_{b_j \in \mathcal{B}\left(I'_j, B'_j\right)} \left\{ C_2\left(s'_j, b_j\right) + v_{n,j}\left(s''_j\right) \right\}$$

        where: s''_j is from equation (5.13) and (5.14)

      end

      for $s_j = (t, e = 1, I_j, Q_j, B_j = 0, A = 0) \in \{(t, e = 1, I_j, Q_j, B_j = 0, A = 0) \mid I_j \in \mathcal{I}_j, Q_j \in \mathcal{Q}_j(t, I_j)\}$ **do**

$$v_{n,j}(s_j) = \max_{q_j \in \mathcal{Q}_j(t, I_j)} \left\{ C_1(s_j, q_j) + \sum_{d_{1,j}=0}^{\widetilde{D}_{1,j}} \sum_{d_{2,j}=0}^{\widetilde{D}_{2,j}} \sum_{d_0=0}^{\widetilde{D}_0} \widetilde{P}_{1,j}(d_{1,j}) \widetilde{P}_{2,j}(d_{2,j}) \widetilde{P}_0(d_0) \cdot v_{n,j}\left(s'_j\right) \right\}$$

        where: s'_j is from equation (5.9), (5.10), (5.11), and (5.12)

      end

    end

  end

end

Appendix 5.F Value iteration for the OFC

Algorithm 7: Value iteration for the OFC.

initialisation: $v_{0,0} = 0$

for $n = 1 : 1 : n_{max}$ **do**

  for $t = 7 : -1 : 1$ **do**

    for $s_0'' = (t, e = 3, I_0'', Q_0'', B_0'' = 0, A'') \in$
    $\{(t, e = 3, I_0'', Q_0'', B_0'' = 0, A'') \mid I_0'' \in \mathcal{I}_0, Q_0'' \in \mathcal{Q}_0(t, I_0''), A'' \in \mathcal{D}_0\}$ **do**

      $$v_{n,0}(s_0'') = \max_{a_0 \in \mathcal{A}(I_0'', A'')} \left\{ C_3(s_0'', a_0) + v_{n-1,0}(s_0''') \right\}$$ where: s_0''' is from equation (5.15) and (5.16)

    end

    for $s_0' = (t, e = 2, I_0', Q_0', B_0', A') \in$
    $\left\{(t, e = 2, I_0', Q_0', B_0', A') \mid I_0' \in \mathcal{I}_0, Q_0' \in \mathcal{Q}_0(t, I_0'), B_0' \in \widetilde{\mathcal{D}}_{2,0}, A' \in \mathcal{D}_0\right\}$ **do**

      $$v_{n,0}(s_0') = \max_{b_0 \in \mathcal{B}(I_0', B_0')} \left\{ (p - (u_0 + k)) \cdot b_0 - g(B_0' - b_0) + v_{n,0}(s_0'') \right\}$$ where: s_j'' is from equation (5.13) and $B_0'' = 0$

    end

    for $s_0 = (t, e = 1, I_0, Q_0, B_0 = 0, A = 0) \in$
    $\{(t, e = 1, I_0, Q_0, B_0 = 0, A = 0) \mid I_0 \in \mathcal{I}_0, Q_0 \in \mathcal{Q}_0(t, I_0)\}$ **do**

      $$v_{n,0}(s_0) = \max_{q_0 \in \mathcal{Q}_0(t, I_0)} \left\{ C_0(s_0, q_0) + \sum_{d_{2,0}=0}^{\widetilde{D}_{2,0}} \sum_{d_0=0}^{\widetilde{D}_0} \widetilde{P}_{2,0}(d_{2,0}) \widetilde{P}_0(d_0) \cdot v_{n,0}(s_0') \right\}$$
      where: s_j' is from equation (5.9), (5.10), (5.12), and $B_0' = d_{2,0}$

    end

  end

end

Appendix 5.G State transition

State transitions during epoch 1

The state transition for the inventory at each store during epoch 1 is dependent on the amount of in-store demand that is being fulfilled. The inventory state transition is as follow:

$$I'_j = (I_j - d_{1,j})^+ \qquad \forall j \in \{1, 2, \ldots, J\}. \tag{5.9}$$

As the OFC does not receive any in-store demand $I'_0 = I_0$. The transition of outstanding orders is dependent on whether the lead time has expired or not, as we do not need to keep track of the outstanding order after lead time. The transition is as follow:

$$Q'_j = \begin{cases} q_j & \text{if } t = 1 \\ Q_j & \text{if } 1 < t \leq L \\ 0 & \text{else} \end{cases}. \tag{5.10}$$

Furthermore, the accepted BOPS orders of the day is transitioned as follow:

$$B'_j = d_{2,j} \qquad \forall j \in \{1, 2, \ldots, J\}. \tag{5.11}$$

Lastly, the accepted BODH demand of the day is transitioned as follow:

$$A' = d_0. \tag{5.12}$$

State transitions during epoch 2

The state transition for the inventory at each store during epoch 2 is dependent on the amount of BOPS orders being fulfilled by the OFC and store. The inventory state transition is as follow:

$$I''_j = I'_j - b_j \qquad \forall j \in \mathcal{J}. \tag{5.13}$$

The transition of outstanding orders does not change during the epoch thus, $Q'' = Q'$. As the accepted BOPS orders of the day are either fulfilled or cancelled the state is set to zero:

$$B''_j = 0 \qquad \forall j \in \{1, 2, \ldots, J\}. \tag{5.14}$$

Lastly, the accepted BODH demand of the day does not change during the epoch as the BODH orders are not fulfilled in this epoch, thus $A'' = A'$.

State transitions during epoch 3

The state transition for the inventory at each store and OFC during epoch 3 is dependent on the amount of BODH orders being fulfilled by the store and if replenishment occurs. The inventory state transition is as follow:

$$I_j''' = I_j'' - a_j + \delta\left(t = L\right) \cdot Q_j'' \qquad \forall j \in \mathcal{J}. \tag{5.15}$$

Here, $\delta(x)$ denotes the Kronecker delta, which gives the value 1 if $x =$ True, otherwise 0. The transition of outstanding orders does not change during the epoch thus, $Q''' = Q''$. Similarly, the accepted BOPS orders of the day does not change during the epoch as no new orders are received, thus $B''' = B''$.

The accepted BODH orders of the day are either fulfilled or cancelled thus the state is set to zero:

$$A''' = 0. \tag{5.16}$$

At the end of the sub-period, the state t is also increased by one, at the end of the period (when $t = R$) then it is set back to one.

Chapter 6

Conclusions

6.1 Findings

This thesis studied the inventory management of an omni-channel retailer that leverages the store inventory for the fulfilment of online orders. Four studies were presented in Chapter 2 to 5. Although these studies all had a different aim and scope, they all investigated the operational decision-making of an omni-channel retailer with single or multiple brick-and-mortar stores. This topic is especially relevant nowadays given the rise in omnichannel retail and the associated challenges in managing omnichannel order fulfilment (Melacini et al., 2018).

We first described how to ration store inventory across in-store and online demand for an omni-channel retailer operating a single store. We then discussed how product returns influence the rationing and replenishment decision. Subsequently, we studied how an omni-channel retailer that uses its network of stores for the fulfilment of online orders can best allocate these to the different stores. Finally, we investigated the performance of different omnichannel fulfilment strategies such as ship-to-store, buy-online-pickup-store, and ship-from-store. To achieve this, we developed Markov Decision Process (MDP) models that capture the characteristics of an omni-channel retailer, which were solved with various methods. We conclude with this chapter, in which we discuss and summarise the main findings from the four studies, their scientific contributions, as well as managerial insights. The chapter ends with future avenues for research.

We first discuss the findings for omni-channel retailers with a single store, followed by the findings of a retailer operating multiple stores. A summary of the main findings of the chapters is given in Table 6.1.

6.1.1 Omni-channel retailers single store operations

When using store inventory for the fulfilment of online orders, the operations of the store need to be revised (Hübner et al., 2022). In managerial studies (e.g. ENC, 2016), it is often suggested to reserve part of the inventory for the online channel. This rationing decision ensures that there is no conflict between the in-store and online customers for the fulfilment of demand. Practically, rationing inventory relates to storing part of the inventory in the backroom to satisfy online demand. Physically separating the products for the two different demands, although keeping the products at the same location, not only has the advantage of reducing conflicts between demands, but it

Table 6.1: Summary of the main findings of this thesis.

Chapter	Research scope	Main findings
2	How can store inventory be utilised for online demand?	• Rationing the inventory is a solution to the negative effects that occur when using in-store inventory for fulfilling online orders • For the rationing policy, a maximum threshold level for high inventory levels is found, at low inventory levels there is a trade-off based on expected demand and costs • The optimal ordering policy of the retailer consists of a fixed order quantity at low inventory levels and an order-up-to level at high inventory levels
3	How is the omni-channel retailer's performance influenced by online returns?	• Modelling the returns as multi-period sales-dependent leads to higher profit and service levels compared to other approaches for modelling returns • Higher online returns leads to lower profit and lower service levels in the offline channel • Longer return windows do not influence the retailer's profit
4	Which store should fulfil an online order in an omni-channel retailers network?	• A fast approximate method is developed based on an exact solution approach to solve the allocation of online orders to stores • Our method has the advantage of making a well-informed online fulfilment decision as it is able to capture the trade-off between fulfilling an online order or reserving the inventory for future demand • Our method outperforms two existing methods on profit and service levels, where our method is better at balancing inventories across the stores in the network

Continued on the next page

Table 6.1: Summary of the main findings of this thesis (continued).

Chapter	Research scope	Main findings
5	What is the best fulfilment strategy for an omni-channel retailers network?	• Including the stores in online demand fulfilment results in higher overall service levels • Allowing both stores and OFC to fulfil demand is most profitable, but only using stores will result in higher overall service level • The added value of using the store or OFC for the fulfilment of orders placed online is dependent on the inventory cost and order fulfilment cost differences between the locations

can also reduce inventory cost, as fewer products are displayed in-store which has expensive shelf space (Hübner et al., 2016b). In Chapter 2, it was found that there exists an optimal threshold of the number of products to be stored in-store when the inventory level is high enough to satisfy most demand. This threshold is based on the trade-off between the economic indicators and expected demand.

In addition to rationing the inventory, the retailer also needs to regularly replenish the store inventory. The ordering decision normally occurs less regular than the rationing decision (e.g., weekly instead of daily) and the resulting inventory replenishment typically happens after a fixed lead time. In Chapter 2, an optimal ordering policy was identified, consisting of a fixed order quantity at low inventory levels and an order-up-to-level at high inventory levels. The structure of this policy is due to the uncertainty of demand during the procurement lead time. When the retailer uses the store inventory for both in-store demand and online demand, an inventory pooling effect is created, which has the advantage of reducing uncertainty in demand resulting in lower costs to hold safety stock.

Selling products online can improve overall sales for the omni-channel retailer. However, products that are ordered online are often being returned (de Leeuw et al., 2016). Retailers prefer customers to return products in-store as it is more cost-efficient and it allows the customer to get a direct refund or replacement (Wollenburg et al., 2018;

Nageswaran et al., 2020). Therefore, in Chapter 3, we extend the study in Chapter 2 by investigating the influence of product returns on the omni-channel retailer's performance. The uncertainty in the quantity and timing of returned products negatively influences the retailer's inventory control, as it can cause excessive stock (Hu et al., 2019). To study the problem, we explicitly model multi-period sales-dependent returns, which is more realistic in contrast to previous literature. but increases the complexity of the studied problem. We found that taking this information into account is of importance, as aggregating or simplifying the information about returns results in lower profits and service levels.

However, due to the two channels being managed from the same inventory, negative characteristics from one channel can influence the other. The online returns reduce the profitability of selling products online. With lower profit per sold products the order quantities will be reduced which results in lower in-store inventory levels. The results show that a high return rate negatively influences the in-store service level and profit due to low product profitability. However, longer return windows do not influence the retailers' profit as it doesn't affect the profitability of a product. The profitability of a product is dependent on the expected revenue and costs which does not depend on the return window, as the return cost depend on whether a product is returned rather than when.

6.1.2 Omni-channel retailers network operations

When the omni-channel retailer has multiple stores available, online orders need to be allocated to the stores. This is however not straightforward, as the retailer has to take into account all store inventory levels, outstanding orders, and expected future demands. As found in Chapter 4, if the retailer does not take into account all this information, it can result in inventory levels becoming imbalanced across the retailer's network. This can result in excessive stocks for some stores while other stores face stock-outs. By formulating the problem as an MDP, this information can be taken into account. As such a model easily becomes intractable with multiple stores, we develop an approximation of the optimal policy by decomposing the MDP and applying a one-step policy improvement approach. It was discovered that this approach outperformed well-known heuristics as it better allocates the orders to stores to keep the inventories balanced.

Moreover, when demand is fulfilled from stores, the store inventory needs to be increased. This has the added advantage that the service level for the in-store customer is increased. Also for the online channel higher service rates is achieved as demand is not limited to one stock, as now there are multiple locations available for fulfilling an online order.

The added value and role of an online fulfilment centre (OFC) in an omni-channel retail network were investigated in Chapter 5. Hübner et al. (2022) mentions four typical fulfilment strategies which we investigated with different roles in the fulfilment of demand for the stores and OFC. It was confirmed that allowing both stores and OFC for fulfilling all orders that are placed online is most profitable, however only using stores results in higher overall service levels. The added value of using the store or OFC for the fulfilment of orders placed online is dependent on the fulfilment cost difference between the locations; with low differences between costs only using stores is preferred.

6.2 Scientific Contributions

The presented work in this thesis contributes to the existing literature on omni-channel retail operations in several ways.

First, in all chapters we focus on the store inventory management of an omni-channel retailer. Although the literature does discuss the role of the store in omni-channel retailing (e.g. Govindarajan et al., 2021; Pichka et al., 2022), most of the literature only focuses on a single decision being taken by the retailer, while with omni-channel retail operations often the retailer has multiple decisions to be made at different time periods. One of the key decisions to be made is the replenishment decision, which is often investigated in the literature but only for simple settings with finite selling horizons (e.g. Jia et al., 2021). Additionally, the replenishment is often considered to be instantaneous while in a more realistic setting there is a delivery lead time. In our study, we do include lead times, and a non-finite selling horizon. Additionally, we investigate different types of decisions that omni-channel retailers can make and model these across different time periods, as some processes (e.g., rationing and online order fulfilment) occur more regularly than others (e.g., inventory replenishment).

Second, in Chapter 3 we looked at how online returns might influence the omni-channel retailer. Contrary to previous studies, the returns in our study are modelled multi-period and sales-dependent, which increases the model complexity but makes it more realistic (Benedito & Corominas, 2013; Ambilkar et al., 2022). We deal with the complexity by using reinforcement learning to find a good performing policy.

Third, in Chapter 5, we included novel concepts and characteristics of omni-channel retailers in our studies such as ship-from-store, ship-to-store, and buy-online-pick-up-in-store strategies. Whereas most of the literature only focuses on a single concept (e.g. Arslan et al., 2021; Govindarajan et al., 2021), we investigated the inventory management when the omni-channel retailer implements multiple.

Fourth, we applied the conventional framework of MDP for solving these contemporary problems. The advantage of using MDP is that it takes into account all information of the current system and potential future outcomes. It is observed from the studies in this thesis that taking such information into account has a clear added value on performance in terms of profit and service levels. In Chapter 3, we observed that other methods that aggregate states or approximate state transitions to reduce the complexity of the MDP results in information loss and therefore also lower performance. Additionally, in Chapter 4 it was observed that myopic rules that do not take into account potential future outcomes were outperformed by our approach. Our approach was found via an approximate solution of the MDP that makes a well-balanced decision between short and far-sighted outcomes.

Lastly, as some problems became too large we used more novel techniques such as Deep Reinforcement Learning (Chapter 3) and one-step policy improvement (Chapter 4). We showed that these solution approaches are capable of finding good-performing policies for the types of problems we study. Another solution to circumvent the issue of problems becoming too large is to derive heuristics from the optimal policy of smaller instances. In Chapter 2, we developed heuristics for the rationing and ordering decisions, which showed near-optimal results. The advantage of such heuristics is that they can scale to multiple products and stores, and that they are easier to implement in practice.

6.3 Managerial Insights

The current retail industry has changed drastically in the last few decades with the rise of online shopping and redefining the role of brick-and-mortar stores. This thesis provides insights into the inventory management of omni-channel retailers who want to use their store inventory for the online channel. Several managerial insights were derived on the operational decision-making behind omni-channel retailing.

Firstly, to mitigate leveraging the assets of the brick-and-mortar store for the online channel would negatively impact the in-store customers, retailers need to apply proper inventory management. One of the expected negative side effects was that in-store customers might experience stock-outs due to products being used to fulfil online demand. However, as seen in this thesis, with proper inventory management, an actual synergy instead of conflict can arise between the two channels, resulting in higher service levels for all customers. This synergy arises due to the demand pooling effect that is created at the store. The in-store customers benefit the most from the higher stock levels at the stores when it needs to fulfil multiple demands. In-store demand is often prioritised as they are the most profitable due to requiring no extra costs such as shipping or potentially handling returns.

Secondly, to avoid possible conflicts arising between the online customers and in-store customers, retailers should reserve part of the in-store inventory for online customers. This can be done by storing part of the inventory in the backroom. When done properly, the advantage of demand pooling can still be created without influencing the customer's shopping experience or the retailer's profit. To decide on the number of products to reserve for a channel, the retailer should take into account the cost structure of each channel and their respective expected demands.

Thirdly, when the omni-channel retailer is experiencing products being returned from the online sales channel, they should take these into account in their inventory management. More specifically, they should take into account the quantity and timing of products being sold online. With this information, the timing of when products are being returned can be better predicted and be used in the management of their inventory.

Fourthly, the retailer should allow flexible fulfilment of demand. By allowing demand to be fulfilled from any location and thus not limiting the fulfilment to dedicated locations, higher profit and service levels can be achieved. For example, when an omni-channel retailer has dedicated OFCs available, they should still include stores in the fulfilment of online demand as this can result in better performance.

Lastly, when deciding from which location to fulfil an online order it is important to take into account all store inventory levels, outstanding orders, and expected future demand. When not taking into account this information (and making a myopic decision), the store inventory levels can become imbalanced across the network, resulting in stock-outs at certain stores while others might have excessive stock. Thus, the retailer should carefully allocate the orders and not base the allocation decision only on for example the store inventory levels.

6.4 Further Research

Based on the findings from this thesis, this section discusses several suggestions for future research. We first discuss more methodological research directions related to the modelling of inventory management. Secondly, we describe research directions related to the problem context of retail store inventory management in an omni-channel context. These suggestions are not only limited to extending the current research but also to a broader range of topics related to modelling of inventory management and omni-channel retailing.

6.4.1 Modelling inventory management

Markov Decision Processes are a suitable framework for modelling decision-making in a stochastic environment. However, they have some limitations in their applicability. To find the optimal policy, it needs to iterate over each possible state-action combination and potential outcomes multiple times. As the models can easily become too large with multi-dimensional states, the MDP can become intractable. Therefore often approximate dynamic programming is applied, which are algorithms such as reinforcement learning that try to overcome the curse of dimensionality.

The application of reinforcement learning in inventory management settings does however have some limitations. Reinforcement learning has been developed with more of a focus on deterministic settings: if an action is taken, the outcome is considered to be fixed. As demand is often modelled as stochastic in inventory management, reinforcement learning is hindered in the performance of finding a good-performing policy. Therefore, an interesting research direction is to improve reinforcement learning algorithms to perform better under uncertainty.

Additionally, reinforcement learning learns about the problem without any prior knowledge. However, in inventory management often a good initial policy can be found from heuristics that study similar topics. Therefore, using these heuristics as an initial start for reinforcement learning could reduce training time and potentially allow for finding better policies faster. De Moor et al. (2022) show how rewards can be reshaped with well-known heuristics to accelerate training and learning; however, their algorithm still starts with no initial start. For instance, the effect of transfer learning, where the policy of a heuristic could be used for the learning is of interest.

Lastly, as mentioned in Gijsbrechts et al. (2022), reinforcement learning is a black box and gives little insight in the structure of the found policy and how it was derived. This might hinder the application in practice, as lack of interpretability can undermine the trust in the policy.

6.4.2 Omni-channel retailing

This thesis only focused on the inventory management of a single product, while customers might place online orders consisting of multiple items. Therefore, the allocation of online orders to stores and OFCs becomes more complicated as an extra decision needs to be made if the order should be fulfilled from a single location or should be split over multiple locations. This decision does not only influence the store inventory management and potentially increase the shipping cost, but it might also influence customer satisfaction as receiving multiple shipments might not be preferred. Therefore, from both a marketing and operations perspective, this decision needs to be optimised.

Additionally, the omni-channel retailer has extra inventory operations that could be implemented to improve performance. The retailer could decide to transship products between the stores and/or OFCs to rebalance the inventory across the retailer's network, which was not considered in this thesis. Especially when considering product returns, transhipment might be needed, as this can result in excessive stock at one location due to less inventory control. Dijkstra et al. (2019) study such a setting; however, their study is limited to dual-channel and not omni-channel retailers. Their findings do show promising results with regards to keeping inventories balanced across the network, which potentially translates well to the omni-channel context.

Another decision that the retailer could implement is to backlog online orders. Although customers often expect next-day delivery, it is not uncommon that delivery can take multiple days. By backlogging online orders, retailers can keep their current inventory available for other customers. However, delaying the delivery of online orders might result in lower customer satisfaction, thus a careful decision needs to be made to ensure overall high customer satisfaction.

In this thesis, it is assumed that demand is stationary and non-seasonal. Although this is a common assumption in inventory management, it does not always hold in practice (Hübner et al., 2022). Some products are seasonal (e.g., fashion) and thus their inventory management might differ, for instance there is no replenishment option and a fixed selling horizon. The studies in this thesis can easily be adapted to a fixed selling horizon with a single replenishment option, and it would even reduce the computational complexity. It would be interesting to see what the performance of the policies found in this thesis is when we relax the assumption of demand being stationary and non-seasonal. Additional investigation could be performed in if the policy found for our multi-period problem can be extended to be used for single period problems.

In our thesis we assume that when customers encounter a stock-out the demand is lost. However, customers might switch to the channel which does hold inventory, therefore they could be steered to visit another channel. This could influence the inventory management of the retailer; however, the magnitude of the effect needs to be measured, as it might unnecessarily complicate decision-making when the benefits are small.

Finally, this thesis has focused primarily on brick-and-mortar stores integrating an online channel and leveraging the store inventory for the fulfilment of online orders. However, online retailers are becoming increasingly present in the physical spaces, not only by opening stores but also with the introduction of zero-inventory showrooms. They use such showrooms to showcase their products, but the customer cannot fulfil their needs in-store (Bell et al., 2018a). Instead, the retailer offers the customer to buy the product digitally (followed by in-store pickup or home delivery). An intermediate alternative between showrooms and brick-and-mortar stores is digital assortment extension, in which digital devices in the store enable the retailer to expand their assortment as online channels can more easily hold more stock. For these new concepts, the retailer has to decide the assortment to be displayed, deciding which products the customers would like to see in the showroom. Some products might influence the customers perceived value of similar products. For example, when one fashion product is displayed in the size of the customer their uncertainty about the sizing of similar fashion products is removed.

As more retailers adopt omni-channel retail operations, the related challenges will remain an important topic to research. The list of future research topics is not limited to the ones presented in this chapter. As mentioned in Rooderkerk et al. (2023), the increase in data availability in omni-channel retailing, together with researchers' knowledge, could further improve retail operations while maintaining customer satisfaction.

References

Aastrup, J., & Kotzab, H. (2010). Forty years of out-of-stock research-and shelves are still empty. *The International Review of Retail, Distribution and Consumer Research*, *20*, 147–164.

Abdulla, H., Ketzenberg, M., & Abbey, J. D. (2019). Taking stock of consumer returns: A review and classification of the literature. *Journal of Operations Management*, *65*, 560–605.

Abrudan, I. N., Dabija, D.-C., & Grant, D. B. (2020). Omni-channel retailing strategy and research agenda. In *Perspectives on Consumer Behaviour: Theoretical Aspects and Practical Applications* (pp. 261–280).

Acimovic, J., & Graves, S. C. (2015). Making better fulfillment decisions on the fly in an online retail environment. *Manufacturing & Service Operations Management*, *17*, 34–51.

Adan, I., van Eenige, M., & Resing, J. (1995). Fitting discrete distributions on the first two moments. *Probability in the Engineering and Informational Sciences*, *9*, 623–632.

Agatz, N. A., Fleischmann, M., & Van Nunen, J. A. (2008). E-fulfillment and multi-channel distribution - A review. *European Journal of Operational Research*, *187*, 339–356.

Ailawadi, K. L., & Farris, P. W. (2017). Managing multi- and omni-channel distribution: Metrics and research directions. *Journal of Retailing*, *93*, 120–135.

Aksen, D., & Altinkemer, K. (2008). A location-routing problem for the conversion to the "click-and-mortar" retailing: The static case. *European Journal of Operational Research*, *186*, 554–575.

Akturk, M. S., Ketzenberg, M., & Heim, G. R. (2018). Assessing impacts of introducing ship-to-store service on sales and returns in omnichannel retailing: A data analytics study. *Journal of Operations Management*, *61*, 15–45.

Alfieri, A., Pastore, E., & Zotteri, G. (2017). Dynamic inventory rationing: How to allocate stock according to managerial priorities. An empirical study. *International Journal of Production Economics*, *189*, 14–29.

Ambilkar, P., Dohale, V., Gunasekaran, A., & Bilolikar, V. (2022). Product returns management: A comprehensive review and future research agenda. *International Journal of Production Research*, *60*, 3920–3944.

Andrews, J. M., Farias, V. F., Khojandi, A. I., & Yan, C. M. (2019). Primal–dual algorithms for order fulfillment at Urban Outfitters, Inc. *INFORMS Journal on Applied Analytics*, *49*, 355–370.

Arslan, A. N., Klibi, W., & Montreuil, B. (2021). Distribution network deployment for omnichannel retailing. *European Journal of Operational Research*, *294*, 1042–1058.

Baird, N., & Kilcourse, B. (2011). *Omni-channel fulfillment and the future of the retail supply chain*. Benchmark Report, Retail Systems Research.

Bao, L., Liu, Z., Yu, Y., & Zhang, W. (2018). On the decomposition property for a dynamic inventory rationing problem with multiple demand classes and backorder. *European Journal of Operational Research*, *265*, 99–106.

Bayram, A., & Cesaret, B. (2021). Order fulfillment policies for ship-from-store implementation in omni-channel retailing. *European Journal of Operational Research*, *294*, 987–1002.

Bell, D. R., Gallino, S., & Moreno, A. (2014). How to win in an omnichannel world. *MIT Sloan Management Review*, *56*, 45 – 53.

Bell, D. R., Gallino, S., & Moreno, A. (2018a). Offline showrooms in omnichannel retail: Demand and operational benefits. *Management Science*, *64*, 1629–1651.

Bell, D. R., Gallino, S., & Moreno, A. (2018b). The store is dead-long live the store. *MIT Sloan Management Review*, *59*, 59–66.

Bendoly, E. (2004). Integrated inventory pooling for firms servicing both on-line and store demand. *Computers & Operations Research*, *31*, 1465–1480.

Bendoly, E., Blocher, D., Bretthauer, K. M., & Venkataramanan, M. (2007). Service and cost benefits through clicks-and-mortar integration: Implications for the centralization/decentralization debate. *European Journal of Operational Research*, *180*, 426–442.

Benedito, E., & Corominas, A. (2013). Optimal manufacturing policy in a reverse logistic system with dependent stochastic returns and limited capacities. *International Journal of Production Research*, *51*, 189–201.

Benjaafar, S., ElHafsi, M., & Huang, T. (2010). Optimal control of a production-inventory system with both backorders and lost sales. *Naval Research Logistics*, *57*, 252–265.

Bernon, M., Cullen, J., & Gorst, J. (2016). Online retail returns management: Integration within an omni-channel distribution context. *International Journal of Physical Distribution & Logistics Management*, *46*, 584–605.

Bhatnagar, R., & Lin, B. (2019). The joint transshipment and production control policies for multi-location production/inventory systems. *European Journal of Operational Research*, *275*, 957–970.

Bhulai, S. (2009). Dynamic routing policies for multiskill call centers. *Probability in the Engineering and Informational Sciences*, *23*, 101–119.

Bijmolt, T. H., Broekhuis, M., de Leeuw, S., Hirche, C., Rooderkerk, R. P., Sousa, R., & Zhu, S. X. (2021). Challenges at the marketing–operations interface in omni-channel retail environments. *Journal of Business Research*, *122*, 864–874.

Boute, R. N., Gijsbrechts, J., van Jaarsveld, W., & Vanvuchelen, N. (2022). Deep reinforcement learning for inventory control: A roadmap. *European Journal of Operational Research*, *298*, 401–412.

Bretthauer, K. M., Mahar, S., & Venakataramanan, M. (2010). Inventory and distribution strategies for retail/e-tail organizations. *Computers & Industrial Engineering*, *58*, 119–132.

de Brito, M. P., & Dekker, R. (2003). Modelling product returns in inventory control - Exploring the validity of general assumptions. *International Journal of Production Economics*, *81*, 225–241.

Brynjolfsson, E., Hu, Y. J., & Rahman, M. S. (2013). Competing in the age of omnichannel retailing. *MIT Sloan Management Review*, *554*, 1–7.

Carmona, M. (2022). The existential crisis of traditional shopping streets: The sun model and the place attraction paradigm. *Journal of Urban Design*, *27*, 1–35.

Caro, F., Kök, A. G., & Martínez-de Albéniz, V. (2020). The future of retail operations. *Manufacturing & Service Operations Management*, *22*, 47–58.

Carr, S., & Duenyas, I. (2000). Optimal admission control and sequencing in a make-to-stock/make-to-order production system. *Operations Research*, *48*, 709–720.

Chen, J., & Bell, P. C. (2012). Implementing market segmentation using full-refund and no-refund customer returns policies in a dual-channel supply chain structure. *International Journal of Production Economics*, *136*, 56–66.

Chew, E. P., Lee, L. H., & Liu, S. (2013). Dynamic rationing and ordering policies for multiple demand classes. *OR Spectrum*, *35*, 127–151.

Chi, C.-Y., Tsai, R. T.-H., Lai, J.-Y., & Hsu, J. Y.-j. (2010). A reinforcement learning approach to emotion-based automatic playlist generation. In *2010 International Conference on Technologies and Applications of Artificial Intelligence* (pp. 60–65). IEEE.

Cohen, A. C. (1954). Estimation of the poisson parameter from truncated samples and from censored samples. *Journal of the American Statistical Association*, *49*, 158–168.

Cui, T. H., Ghose, A., Halaburda, H., Iyengar, R., Pauwels, K., Sriram, S., Tucker, C., & Venkataraman, S. (2021). Informational challenges in omnichannel marketing: Remedies and future research. *Journal of Marketing*, *85*, 103–120.

De Moor, B. J., Gijsbrechts, J., & Boute, R. N. (2022). Reward shaping to improve the performance of deep reinforcement learning in perishable inventory management. *European Journal of Operational Research*, *301*, 535–545.

DeCroix, G. A. (2006). Optimal policy for a multiechelon inventory system with remanufacturing. *Operations Research*, *54*, 532–543.

Deloitte (2015). *Omni-channel retail - A Deloitte point of view*.

Deloitte (2020). *Omni-channel point of view 5 product returns: Closing the omni-channel loop*.

Deshpande, V., Cohen, M. A., & Donohue, K. (2003). A threshold inventory rationing policy for service-differentiated demand classes. *Management Science*, *49*, 683–703.

Dethlefs, C., Ostermeier, M., & Hübner, A. (2022). Rapid fulfillment of online orders in omnichannel grocery retailing. *EURO Journal on Transportation and Logistics*, *11*, 100082.

Difrancesco, R. M., van Schilt, I. M., & Winkenbach, M. (2021). Optimal in-store fulfillment policies for online orders in an omni-channel retail environment. *European Journal of Operational Research*, *293*, 1058–1076.

Dijkstra, A. S., Van der Heide, G., & Roodbergen, K. J. (2019). Transshipments of cross-channel returned products. *International Journal of Production Economics*, *209*, 70–77.

ENC (2016). What every omni-channel retailer should know about developing a ship from store strategy. enVista Company.

EY (2015). *Re-engineering the supply chain for the omni-channel of tomorrow.*

Fadıloğlu, M. M., & Bulut, Ö. (2010). A dynamic rationing policy for continuous-review inventory systems. *European Journal of Operational Research*, *202*, 675–685.

Feinberg, E. A., & Lewis, M. E. (2005). Optimality of four-threshold policies in inventory systems with customer returns and borrowing/storage options. *Probability in the Engineering and Informational Sciences*, *19*, 45–71.

Fernie, J., & Sparks, L. (2004). *Logistics and retail management: Insights into current practice and trends from leading experts*. Kogan Page Publishers.

Fleischmann, M., & Kuik, R. (2003). On optimal inventory control with independent stochastic item returns. *European Journal of Operational Research*, *151*, 25–37.

Fleischmann, M., Kuik, R., & Dekker, R. (2002). Controlling inventories with stochastic item returns: A basic model. *European Journal of Operational Research*, *138*, 63–75.

Fleischmann, M., & Minner, S. (2004). Inventory management in closed loop supply chains. In *Supply Chain Management and Reverse Logistics* (pp. 115–138). Springer.

Frank, K. C., Zhang, R. Q., & Duenyas, I. (2003). Optimal policies for inventory systems with priority demand classes. *Operations Research*, *51*, 993–1002.

Gao, F., Agrawal, V. V., & Cui, S. (2022). The effect of multichannel and omnichannel retailing on physical stores. *Management Science*, *68*, 809–826.

Gao, F., & Su, X. (2017a). Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science*, *63*, 2478–2492.

Gao, F., & Su, X. (2017b). Online and offline information for omnichannel retailing. *Manufacturing & Service Operations Management*, *19*, 84–98.

Gayon, J. P., De Vericourt, F., & Karaesmen, F. (2009). Stock rationing in an M/E r/1 multi-class make-to-stock queue with backorders. *IIE Transactions*, *41*, 1096–1109.

Gijsbrechts, J., Boute, R. N., Van Mieghem, J. A., & Zhang, D. (2022). Can deep reinforcement learning improve inventory management? Performance on dual sourcing, lost sales and multi-echelon problems. *Manufacturing & Service Operations Management*, *24*, 1349–1368.

Goedhart, J., Haijema, R., & Akkerman, R. (2022). Inventory rationing and replenishment for an omni-channel retailer. *Computers & Operations Research*, *140*, 105647.

Goedhart, J., Haijema, R., Akkerman, R., & de Leeuw, S. (2023). *Replenishment and fulfilment decisions for stores in an omni-channel retail network*. Working Paper, Wageningen University & Research.

Goraya, M. A. S., Zhu, J., Akram, M. S., Shareef, M. A., Malik, A., & Bhatti, Z. A. (2022). The impact of channel integration on consumers' channel preferences: Do showrooming and webrooming behaviors matter? *Journal of Retailing and Consumer Services*, *65*, 102130.

Govindarajan, A., Sinha, A., & Uichanco, J. (2021). Joint inventory and fulfillment decisions for omnichannel retail networks. *Naval Research Logistics*, *68*, 779–794.

Gupta, V. K., Ting, Q., & Tiwari, M. K. (2019). Multi-period price optimization problem for omnichannel retailers accounting for customer heterogeneity. *International Journal of Production Economics*, *212*, 155–167.

Ha, A. Y. (1997). Inventory rationing in a make-to-stock production system with several demand classes and lost sales. *Management Science*, *43*, 1093–1103.

Haijema, R., & Hendrix, E. M. T. (2014). Traffic responsive control of intersections with predicted arrival times: A Markovian approach. *Computer-Aided Civil and Infrastructure Engineering*, *29*, 123–139.

Haijema, R., van der Wal, J., & van Dijk, N. M. (2007). Blood platelet production: Optimization by dynamic programming and simulation. *Computers & Operations Research*, *34*, 760–779.

Haynsworth, H. C., & Price, B. A. (1989). A model for use in the rationing of inventory during lead time. *Naval Research Logistics*, *36*, 491–506.

Herhausen, D., Binder, J., Schoegel, M., & Herrmann, A. (2015). Integrating bricks with clicks: Retailer-level and channel-level outcomes of online-offline channel integration. *Journal of Retailing*, *91*, 309–325.

Hobkirk, I. (2015). Key distribution strategies of top omni-channel retailers. *Logistics Management*, *55*, 48–53.

Hovelaque, V., Soler, L. G., & Hafsa, S. (2007). Supply chain organization and e-commerce: A model to analyze store-picking, warehouse-picking and drop-shipping. *4OR*, *5*, 143–155.

Hu, M., Xu, X., Xue, W., & Yang, Y. (2022). Demand pooling in omnichannel operations. *Management Science*, *68*, 883–894.

Hu, X., Wan, Z., & Murthy, N. N. (2019). Dynamic pricing of limited inventories with product returns. *Manufacturing & Service Operations Management*, *21*, 501–518.

Huang, S., & Ontañón, S. (2020). A closer look at invalid action masking in policy gradient algorithms. *arXiv preprint arXiv:2006.14171*, .

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*, 73 – 101.

Hübner, A., Hense, J., & Dethlefs, C. (2022). The revival of retail stores via omnichannel operations: A literature review and research framework. *European Journal of Operational Research*, *302*, 799–818.

Hübner, A., Holzapfel, A., & Kuhn, H. (2015). Operations management in multi-channel retailing: An exploratory study. *Operations Management Research*, *8*, 84–100.

Hübner, A., Holzapfel, A., & Kuhn, H. (2016a). Distribution systems in omni-channel retailing. *Business Research*, *9*, 255–296.

Hübner, A., Wollenburg, J., & Holzapfel, A. (2016b). Retail logistics in the transition from multi-channel to omni-channel. *International Journal of Physical Distribution & Logistics Management*, *46*, 562–583.

Hübner, A. H., Kuhn, H., & Wollenburg, J. (2016c). Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework. *International Journal of Retail & Distribution Management*, *44*.

Hung, H. C., Chew, E. P., Lee, L. H., & Liu, S. (2012). Dynamic inventory rationing for systems with multiple demand classes and general demand processes. *International Journal of Production Economics*, *139*, 351–358.

Hung, Y. F., & Hsiao, J. Y. (2013). Inventory rationing decision models during replenishment lead time. *International Journal of Production Economics*, *144*, 290–300.

Ishfaq, R., & Bajwa, N. (2019). Profitability of online order fulfillment in multi-channel retailing. *European Journal of Operational Research*, *272*, 1028–1040.

Ishfaq, R., Defee, C. C., Gibson, B. J., & Raja, U. (2016). Realignment of the physical distribution process in omni-channel fulfillment. *International Journal of Physical Distribution & Logistics Management*, *46*, 543–561.

Ishfaq, R., & Raja, U. (2018). Evaluation of order fulfillment options in retail supply chains. *Decision Sciences*, *49*, 487–521.

Jalilipour Alishah, E., Moinzadeh, K., & Zhou, Y.-P. (2015). *Inventory fulfillment strategies for an omni-channel retailer*. Working Paper, University of Washington.

Janakiraman, N., Syrdal, H. A., & Freling, R. (2016). The effect of return policy leniency on consumer purchase and return decisions: A meta-analytic review. *Journal of Retailing*, *92*, 226–235.

Jasin, S., Sinha, A., & Uichanco, J. (2019). Omnichannel operations: Challenges, opportunities, and models. In *Operations in an Omnichannel World* (pp. 15–34).

Jia, S., Karp, J., Ravi, R., & Tayur, S. (2021). Effective online order acceptance policies for omnichannel fulfillment. *Manufacturing & Service Operations Management*, *24*, 1650–1663.

Jin, D., Caliskan-Demirag, O., Chen, F. Y., & Huang, M. (2020). Omnichannel retailers' return policy strategies in the presence of competition. *International Journal of Production Economics*, *225*, 107595.

Jiu, S. (2022). Robust omnichannel retail operations with the implementation of ship-from-store. *Transportation Research Part E: Logistics and Transportation Review*, *157*, 102550.

Johnson, N. L., Kotz, S., & Kemp, A. W. (2005). *Univariate discrete distributions*. John Wiley & Sons.

Kaplan, A. (1969). Stock rationing. *Management Science*, *15*, 260–267.

Ketzenberg, M. E., Abbey, J. D., Heim, G. R., & Kumar, S. (2020). Assessing customer return behaviors through data analytics. *Journal of Operations Management*, *66*, 622–645.

Kiesmüller, G. P., & Van der Laan, E. A. (2001). An inventory model with dependent product demands and returns. *International Journal of Production Economics*, *72*, 73–87.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, .

Kristensen, A. R. (1988). Hierarchic Markov processes and their applications in replacement models. *European Journal of Operational Research*, *35*, 207–215.

LeCun, Y. A., Bottou, L., Orr, G. B., & Müller, K.-R. (2012). Efficient backprop. In *Neural networks: Tricks of the trade* (pp. 9–48). Springer.

de Leeuw, S., Minguela-Rata, B., Sabet, E., Boter, J., & Sigurðardóttir, R. (2016). Tradeoffs in managing commercial consumer returns for online apparel retail. *International Journal of Operations & Production Management*, *36*, 710–731.

Li, T., Zhao, X., & Xie, J. (2015). Inventory management for dual sales channels with inventory-level-dependent demand. *Journal of the Operational Research Society*, *66*, 488–499.

Li, Z., Wang, J., & Liu, J. (2022). Integrative strategies for omnichannel order fulfillment with risk aversion. *IEEE Transactions on Engineering Management*, .

Lin, Y. H., Wang, Y., Lee, L. H., & Chew, E. P. (2022). Omnichannel facility location and fulfillment optimization. *Transportation Research Part B: Methodological*, *163*, 187–209.

Liu, S., Song, M., Tan, K. C., & Zhang, C. (2015). Multi-class dynamic inventory rationing with stochastic demands and backordering. *European Journal of Operational Research*, *244*, 153–163.

Liu, S., & Zhang, C. (2015). An efficient inventory rationing strategy for systems with two demand classes and backordering. *International Journal of Production Research*, *53*, 6136–6142.

Louchard, G., & Ward, M. D. (2015). The truncated geometric election algorithm: Duration of the election. *Statistics & Probability Letters*, *101*, 40–48.

Ma, S., & Jemai, Z. (2019). Inventory rationing for the news-vendor problem with a drop-shipping option. *Applied Mathematical Modelling*, *71*, 438–451.

Mahar, S., Bretthauer, K. M., & Venkataramanan, M. (2009). The value of virtual pooling in dual sales channel supply chains. *European Journal of Operational Research*, *192*, 561–575.

Mahar, S., Modi, S., & Salzarulo, P. A. (2021). Appreciating how your bread is buttered: Improving online order allocation for cross-channel retailers. *International Journal of Production Research*, *60*, 1–23.

Mahar, S., Salzarulo, P. A., & Wright, P. D. (2012). Using online pickup site inclusion policies to manage demand in retail/e-tail organizations. *Computers & Operations Research*, *39*, 991–999.

Mahar, S., & Wright, P. D. (2009). The value of postponing online fulfillment decisions in multi-channel retail/e-tail organizations. *Computers & Operations Research*, *36*, 3061–3072.

Mandal, P., Basu, P., & Saha, K. (2021). Forays into omnichannel: An online retailer's strategies for managing product returns. *European Journal of Operational Research*, *292*, 633–651.

Marhamat, B. (2022). The history of brick-and-mortar retail and what's in store for the future. *Forbes*.

McKinsey & Company (2020). *Retail optimization*.

McKinsey & Company (2021a). *Retail's need for speed: Unlocking value in omnichannel delivery*.

McKinsey & Company (2021b). *Returning to order: Improving returns management for apparel companies*.

McKinsey & Company (2021c). *The winning formula: What it takes to build leading omnichannel operations*.

Melacini, M., Perotti, S., Rasini, M., & Tappia, E. (2018). E-fulfilment and distribution in omni-channel retailing: A systematic literature review. *International Journal of Physical Distribution & Logistics Management*, *48*, 391–414.

Melchiors, P. (2001). *Rationing policies for an inventory model with several demand classes and stochastic lead times*. Working Paper, Department of Operations Research, University of Aarhus.

Melchiors, P. (2003). Restricted time-remembering policies for the inventory rationing problem. *International Journal of Production Economics*, *81-82*, 461–468.

Melchiors, P., Dekker, R., & Kleijn, M. J. (2000). Inventory rationing in an (s, Q) inventory model with lost sales and two demand classes. *Journal of the Operational Research Society*, *51*, 111–122.

Perea y Monsuwé, T., Dellaert, B. G., & De Ruyter, K. (2004). What drives consumers to shop online? A literature review. *International Journal of Service Industry Management*, *15*, 102–121.

Mou, S., Robb, D. J., & DeHoratius, N. (2018). Retail store operations: Literature review and research directions. *European Journal of Operational Research*, *265*, 399–422.

Muir, W. A., Griffis, S. E., & Whipple, J. M. (2019). A simulation model of multi-echelon retail inventory with cross-channel product returns. *Journal of Business Logistics*, *40*, 322–338.

Nageswaran, L., Cho, S.-H., & Scheller-Wolf, A. (2020). Consumer return policies in omnichannel operations. *Management Science*, *66*, 5558–5575.

Nguyen, D. H., de Leeuw, S., & Dullaert, W. E. (2018). Consumer behaviour and order fulfilment in online retailing: A systematic review. *International Journal of Management Reviews*, *20*, 255–276.

Ni, M., He, Q., Liu, X., & Hampapur, A. (2019). Same-day delivery with crowdshipping and store fulfillment in daily operations. *Transportation Research Procedia*, *38*, 894–913.

Noble, S. M., Griffith, D. A., & Weinberger, M. G. (2005). Consumer derived utilitarian value and channel utilization in a multi-channel retail context. *Journal of Business Research*, *58*, 1643–1651.

Noordhoek, M., Dullaert, W., Lai, D. S., & de Leeuw, S. (2018). A simulation–optimization approach for a service-constrained multi-echelon distribution network. *Transportation Research Part E: Logistics and Transportation Review*, *114*, 292–311.

Omar, H., Klibi, W., Babai, M. Z., & Ducq, Y. (2022). Basket data-driven approach for omnichannel demand forecasting. *International Journal of Production Economics*, (p. 108748).

Ovezmyradov, B., & Kurata, H. (2019). Effects of customer response to fashion product stockout on holding costs, order sizes, and profitability in omnichannel retailing. *International Transactions in Operational Research*, *26*, 200–222.

Pereira, M. M., & Frazzon, E. M. (2021). A data-driven approach to adaptive synchronization of demand and supply in omni-channel retail supply chains. *International Journal of Information Management*, *57*, 102165.

Pichka, K., Alwan, L. C., & Yue, X. (2022). Fulfillment and pricing optimization for omni-channel retailers considering shipment of in-store demand. *Transportation Research Part E: Logistics and Transportation Review*, *167*, 102912.

Piotrowicz, W., & Cuthbertson, R. (2014). Introduction to the special issue information technology in retail: Toward omnichannel retailing. *International Journal of Electronic Commerce*, *18*, 5–16.

Puterman, M. L. (1994). *Markov decision processes*. John Wiley & Sons.

PWC (2020). *How retailers can manage and recover from COVID-19*. Strategy& Retail & Consumer Team.

Radhi, M., & Zhang, G. (2019). Optimal cross-channel return policy in dual-channel retailing systems. *International Journal of Production Economics*, *210*, 184–198.

Rocchetta, R., Bellani, L., Compare, M., Zio, E., & Patelli, E. (2019). A reinforcement learning framework for optimal operation and maintenance of power grids. *Applied Energy*, *241*, 291–301.

Rooderkerk, R. P., de Leeuw, S., & Hübner, A. (2023). Advancing the marketing-operations interface in omnichannel retail. *Journal of Operations Management*, *69*, 188–196.

Rooderkerk, R. P., & Kök, A. G. (2019). Omnichannel assortment planning. In *Operations in an Omnichannel World* (pp. 51–86).

Saha, K., & Bhattacharya, S. (2021). 'buy online and pick up in-store': Implications for the store inventory. *European Journal of Operational Research*, *294*, 906–921.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, .

Shah, S. (1966). On estimating the parameter of a doubly truncated binomial distribution. *Journal of the American Statistical Association*, *61*, 259–263.

Song, P., Wang, Q., Liu, H., & Li, Q. (2020). The value of buy-online-and-pickup-in-store in omni-channel: Evidence from customer usage data. *Production and Operations Management*, *29*, 995–1010.

Tarn, J. M., Razi, M. A., Wen, H. J., & Perez, A. A. (2003). E-fulfillment: The strategy and operational requirements. *Logistics Information Management*, *16*, 350–362.

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

Topkis, D. M. (1968). Optimal ordering and rationing policies in a nonstationary dynamic inventory model with N demand classes. *Management Science*, *15*, 160–176.

Turgay, Z., Karaesmen, F., & Örmeci, E. L. (2015). A dynamic inventory rationing problem with uncertain demand and production rates. *Annals of Operations Research*, *231*, 207–228.

U.S. Department of Commerce (2011). E-commerce 2009.

U.S. Department of Commerce (2021). E-stats 2019: Measuring the electronic economy.

Vanvuchelen, N., Gijsbrechts, J., & Boute, R. (2020). Use of proximal policy optimization for the joint replenishment problem. *Computers in Industry*, *119*, 103239.

Verhoef, P. C., Kannan, P. K., & Inman, J. J. (2015). From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. *Journal of Retailing*, *91*, 174–181.

Wang, D., & Tang, O. (2014). Dynamic inventory rationing with mixed backorders and lost sales. *International Journal of Production Economics*, *149*, 56–67.

Wang, D., Tang, O., & Huo, J. (2013a). A heuristic for rationing inventory in two demand classes with backlog costs and a service constraint. *Computers & Operations Research*, *40*, 2826–2835.

Wang, H., Zhang, W., & He, Y. (2022). Optimal ordering decisions for an omnichannel retailer with ship-to-store and ship-from-store. *International Transactions in Operational Research*, .

Wang, K., Jiang, Z., Li, G., & Cao, J. (2013b). Integrated capacity allocation policies for a production service system with two-class customers and items. In *2013 IEEE International Conference on Automation Science and Engineering (CASE)* (pp. 83–88).

Wiering, M. (2004). Convergence and divergence in standard and averaging reinforcement learning. In *Machine Learning: ECML 2004* (pp. 477–488).

Wijngaard, J. (1979). Decomposition for dynamic programming in production and inventory control. *Engineering and Process Economics*, *4*, 385–388.

Wollenburg, J., Holzapfel, A., Hübner, A., & Kuhn, H. (2018). Configuring retail fulfillment processes for omni-channel customer steering. *International Journal of Electronic Commerce*, *22*, 540–575.

Xiao, Y., Chen, F. Y., & Chen, J. (2009). Optimal inventory and dynamic admission policies for a retailer of seasonal products with affiliate programs and drop-shipping. *Naval Research Logistics*, *56*, 300–317.

Xie, T., Ma, Y., & Wang, Y.-X. (2019). Towards optimal off-policy evaluation for reinforcement learning with marginalized importance sampling. *arXiv preprint arXiv:1906.03393*, .

Xu, J., & Cao, L. (2019). Optimal in-store inventory policy for omnichannel retailers in franchising networks. *International Journal of Retail & Distribution Management*, *47*, 1251–1265.

Xu, X., & Jackson, J. E. (2019). Investigating the influential factors of return channel loyalty in omni-channel retailing. *International Journal of Production Economics*, *216*, 118–132.

Yang, L., Li, X., & Zhong, N. (2022). Omnichannel retail operations with mixed fulfillment strategies. *International Journal of Production Economics*, *254*, 108608.

Zerhouni, H., Gayon, J.-P., & Frein, Y. (2013). Influence of dependency between demands and returns in a reverse logistics system. *International Journal of Production Economics*, *143*, 62–71.

Zhao, N., & Lian, Z. (2011). A queueing-inventory system with two classes of customers. *International Journal of Production Economics*, *129*, 225–231.

Zhou, Y.-W., Zhang, X., Zhong, Y., Cao, B., & Cheng, T. E. (2021). Dynamic pricing and cross-channel fulfillment for omnichannel retailing industry: An approximation policy and implications. *Transportation Research Part E: Logistics and Transportation Review*, *156*, 102524.

Summary

Nowadays, when shopping for products, retail customers are presented with many different options in how you can order and receive a product. Customers can fulfil their demand via online or offline channels and receive the desired product in a store or delivered at a designated location. Enabling customers to fulfil their demands via different integrated shopping channels is referred to as omni-channel retailing. From a marketing perspective, this means that customers do not perceive any difference between the channels when shopping at a retailer. From an operations perspective, this means that the retailer has to coordinate the integration of the channels with a focus on fulfilling the needs of the customer in a cost-efficient manner.

Due to the flexibility of the customer in ordering and receiving their product, a retailer may need to reconsider their retail strategy, as they may need to change their inventory and assortment decisions to different channels in order to satisfy the customer's needs. Especially in relation to the brick-and-mortar store, retailers have to rethink its role. Instead of just a place to shop and buy, a store may be a valuable asset if it can physically support the online channel.

Using the store for the fulfilment of online orders is one of the earliest adopted strategies for leveraging the store assets in omni-channel retail. The advantage of this concept is that it has low setup costs and that it allows for demand to be pooled at the store. However, the disadvantage is that conflict might arise between the channels, as both channels are competing for the same inventory. Therefore, integrating the store assets for the demand of the online channel needs careful operational decision-making, to mitigate the negative side effects without losing the advantages of channel integration.

Therefore, in this thesis, we investigate how the brick-and-mortar store can be used for its new additional roles in omni-channel retailing where it supports the online channel.

We study several key inventory management decisions behind omni-channel retailing and their influence on several performance indicators. This thesis contributes to the current literature on omni-channel retailing by focusing on the store operations. While most of the literature only focuses on a single decision being taken by the retailer, with omni-channel operations often the retailer has multiple decisions to be made at different time periods. Therefore, in this thesis we have taken into account the different types of decisions to be made, e.g. replenishment, rationing, and order allocation. In addition, we investigated the different concepts that omni-channel retailers can introduce for the fulfilment of demand, where we did not focus on one concept but multiple.

In Chapter 2, we investigated how an omni-channel retailer should ration their store inventory to the different demand channels. This rationing decision mitigates conflict between the in-store and online customers for the fulfilment of demand. Practically, it means that some products in the store are reserved for online customers. In addition to a rationing decision, an ordering decision is also needed to replenish the store inventory. To analyse the problem, it was formulated as a Markov Decision Process (MDP) that maximises the expected profit. From the structural properties of the optimal policy, simple yet effective heuristics were derived.

Chapter 3 aimed to investigate a similar problem, but also considers product returns arriving at the store. The uncertainty in the timing and quantity of product returns complicates the replenishment and rationing decisions of the retailer. Contrary to previous studies, we explicitly model the returns as multi-period sales-dependent returns, which is more realistic. We again modelled the problem as an MDP, but as it easily becomes intractable when also considering the product returns, we constructed a Deep Reinforcement Learning (DRL) algorithm to solve the problem. For small-scale instances, we find that the DRL has a low optimality gap; however, its performance drops with a higher level of uncertainty. For large-scale instances, we find that an increase in product returns also has negative side effects for in-store customers, as it leads to the retailer decreasing their order sizes, which increases the risks of stock-outs. However, longer return windows do not influence the retailer's profit.

In Chapter 4, we investigated an omni-channel retailers network, in which there are multiple store inventories available to fulfil an online order. The allocation of online orders to the different stores is complicated, as it needs to take into account potential future outcomes and costs. In practice, the decision is often made myopic, which can result in inventory imbalance across the retailer's stores. We explicitly formulated this problem as a multi-period MDP. To solve this MDP, we decompose it and apply a one-step policy improvement to find a good approximation of the optimal policy. We compared our method with well-known heuristics and found that we outperform these on both profit and service levels. Furthermore, it was shown that our method is better at keeping the store inventories across the network more balanced, resulting in fewer excessive stocks for some stores and fewer stock-outs for other stores.

Chapter 5 is concerned with the decision of how the omni-channel retailer should configure their network for the fulfilment of different types of demand. The customer's demand is taking place either online or offline. When visiting the online channel, the customer is presented with different ways to receive the ordered product: they can have it shipped to their home or collect it in a store. The retailer has to decide for each type of demand whether to use the store inventory of an online fulfilment centre (OFC). We considered four typical network settings used in practice and study the inventory management of the stores and OFC. We modelled the fulfilment and ordering decisions as an MDP and solved it with an adaptation of the method used in Chapter 4. It was found that allowing both stores and OFC fulfilling all orders that are placed online is most profitable. However, only using stores and not an OFC results in higher overall service levels. The added value of using the store or OFC for the fulfilment of orders placed online depends on the fulfilment cost difference between the locations: with low difference between costs only using stores is preferred.

In this thesis, MDP was used as a framework for solving the inventory management decisions of the retailer. As solving the MDP results in the optimal policy, it can be used as a benchmark for other methods to compare performances. We extensively compare our methods with those found in practice or in the literature and can achieve higher profits and service levels. Additionally, the advantage of MDP is that one can identify the structure of the optimal policy and potentially derives simple, yet effective rules from them.

The research presented in this thesis showed the potential of using the store for the operations of an omni-channel retailer. Although previous research mentions that using store assets for the online channel could negatively impact the store customer's shopping experience, this thesis shows that with good inventory management, it could improve the performance of both channels. The demand pooling effect that is created by allocating demand at the store results in a better match between inventory and demand. Furthermore, when multiple stores are available to fulfil the demand of the online channel, it increases the flexibility of the fulfilment decision. This increase in flexibility allows the retailer to decide where to hold inventory as multiple locations can be used for the online demand channel. Not only can it then reduce excessive stock at certain locations, but it can also protect locations with lower inventories from being depleted as it can steer demand fulfilment.

Samenvatting

Wanneer je als klant een product koopt, kan je op veel verschillende manieren je product bestellen en ontvangen. Klanten kunnen hun vraag voor producten via online of offline kanalen vervullen en het product op de door hen gewenste manier ontvangen, in de winkel of thuis. Het klanten in staat stellen hun vraag te voldoen via verschillende kanalen, online en in de winkel, waarbij deze kanalen geïntegreerd zijn, wordt omni-channel retailing genoemd. Vanuit een marketing perspectief betekent dit dat klanten geen verschil ervaren zoals prijs en assortiment tussen de kanalen wanneer zij aan het winkelen zijn. Vanuit operationeel perspectief betekent dit dat de winkelier de integratie van de kanalen moet coördineren met het oog op een kostenefficiënte vervulling van de vraag van de klant.

Vanwege de vele opties voor de klant bij het bestellen en ontvangen van zijn product, moet een winkelier zijn voorraadstrategie heroverwegen, waarbij de voorraad- en assortimentsbeslissingen moet veranderen naar de vraag van de klant. Vooral de rol van de fysieke winkel moet worden heroverwegen: in plaats van een plek om te winkelen kan een winkel een waardevolle ondersteuning bieden voor het online kanaal.

De winkel gebruiken voor het vervullen van online bestellingen is een van de eerste strategieën in omni-channel retailing die wordt toegepast om winkels te betrekken bij het online kanaal. Het voordeel van dit concept is dat de opzetkosten laag zijn en dat de verschillende vragen in de winkel kunnen worden samengevoegd. Het nadeel is echter dat er conflicten kunnen ontstaan tussen de kanalen, aangezien beide kanalen om dezelfde producten concurreren. De integratie van de winkel voor de vraag van het online kanaal vereist daarom zorgvuldige voorraadbeheerstrategieën, om de negatieve neveneffecten te minimaliseren zonder de voordelen van kanaalintegratie te verliezen.

Daarom is er in dit proefschrift gekeken naar hoe de fysieke winkel kan worden gebruikt voor zijn nieuwe extra rollen in omni-channel retailing, waarbij dus de winkel het online kanaal ondersteunt. Wij onderzochten voorraadbeheerstrategieën van een omni-channel winkelier die verschillende prestatie-indicatoren beïnvloeden. Dit proefschrift draagt bij aan de huidige literatuur over omni-channel retailing door zich te richten op de voorraadbeheerstrategieën van winkels. Terwijl de meeste literatuur zich richt op één enkele beslissing van de winkelier, moet wanneer de winkelier omni-channel wordt vaak meerdere beslissingen nemen op verschillende tijdstippen. Daarom hebben wij in dit proefschrift rekening gehouden met de verschillende soorten beslissingen die moeten worden genomen. Daarnaast hebben we de verschillende concepten onderzocht die omni-channel winkeliers kunnen introduceren voor het vervullen van de vraag, waarbij we ons niet op één concept hebben gericht, maar op meerdere.

In hoofdstuk 2 onderzochten we hoe een omni-channel winkelier zijn winkelvoorraad moet verdelen over de verschillende vraagkanalen. De verdeelbeslissing vermindert het conflict tussen de klanten in de winkel en online wanneer zij hun vraag willen voldoen. In de praktijk betekent dit dat een aantal producten achter in de winkel worden opgeslagen voor online klanten. Naast een verdeelbeslissing is ook een bestelbeslissing nodig om de voorraad aan te vullen. Om het probleem te analyseren werd het geformuleerd als een Markov Beslissings Proces (MDP) dat de verwachte winst maximaliseert. Uit de structurele eigenschappen van de optimale beslissingen werden eenvoudige, maar effectieve heuristieken afgeleid.

Hoofdstuk 3 onderzocht een vergelijkbaar bestudeerd probleem als hierboven beschreven, maar introduceerde ook een retourstroom die in de winkel wordt teruggebracht van online bestelde producten. De onzekerheid over de timing en hoeveelheid van productretouren bemoeilijkt de aanvullings- en verdeelbeslissing van de winkelier. In tegenstelling tot eerdere studies hebben wij de retourzendingen expliciet gemodelleerd waarbij we rekening houden met wanneer een online product wordt verkocht. Wij hebben het probleem gemodelleerd als een MDP, maar deze werd al snel niet oplosbaar door de grootte van de toestandsruimte. Daarom hebben wij een Deep Reinforcement Learning-algoritme (DRL) ontwikkeld om het probleem op te lossen. Voor instanties met kleine toestandsruimtes werd ontdekt dat de DRL een lage optimaliteitskloof had, maar dat de prestaties afneemt naarmate de onzekerheid van het probleem toeneemt. Uit grootschalige instanties bleek dat een toename van het aantal productretouren ook negatieve neveneffecten heeft voor klanten in de winkel. Langere retourperiodes hebben echter geen invloed op de winst van de winkelier.

In hoofdstuk 4 onderzochten we een omni-channel retailnetwerk, waarbij er meerdere winkelvoorraden beschikbaar zijn om een online bestelling te vervullen. De toewijzing van een online bestelling aan een winkel is ingewikkeld, omdat er rekening moet worden gehouden met mogelijke toekomstige uitkomsten en kosten. In de praktijk wordt de beslissing vaak kortzichtig genomen, wat goede prestaties oplevert maar kan leiden tot een onevenwichtige voorraden over de winkels in het netwerk. Wij hebben het probleem expliciet geformuleerd als een meerdere periode probleem en opgelost als een periodieke MDP. Het probleem wordt echter onoplosbaar met meerdere winkels, daarom hebben wij de MDP opgebroken in kleinere MDPen die wel oplosbaar zijn. Daarna kunnen we dan via een éénstaps beslissingsverbetering een goede benadering van de optimale beslissing vinden. Wij hebben onze methode vergeleken met bekende heuristieken in de praktijk en vastgesteld dat wij deze overtreffen op zowel winst- als serviceniveau. Er is gebleken dat onze methode beter is in het evenwichtig houden van de winkelvoorraden in het netwerk, wat resulteert in minder overtollige voorraden voor sommige winkels en minder lege schappen voor andere winkels.

Hoofdstuk 5 gaat over de beslissing hoe een omni-channel winkelier zijn netwerk moet configureren om aan verschillende soorten vraag te voldoen. In dit onderzoek hebben wij drie verschillende soorten klanten beschouwd. De vraag van de klant vindt online of offline plaats, waarbij de klant een bestelling plaatst via bijvoorbeeld een webpagina of door een winkel te bezoeken. Bij een bezoek aan het online kanaal krijgt de klant verschillende methoden aangeboden om het bestelde product af te halen: hij kan het thuis laten bezorgen of het online bestelde product in de winkel afhalen. De winkelier moet voor elk type vraag beslissen of hij de winkelvoorraad of een online distributiecentrum te gebruikt. Wij onderzochten vier typische netwerksettings die in de praktijk worden gebruikt, met de nadruk op het voorraadbeheer van de winkels en het distributiecentrum. Om dit proces te analyseren hebben wij de vervul- en bestelbeslissingen gemodelleerd als een MDP en opgelost via de in het vorige hoofdstuk beschreven methode. Het bleek dat het het meest winstgevend is om zowel de winkels als het distributiecentrum te gebruiken om online bestellingen te vervullen, maar dat alleen het gebruik van de winkels leidt tot hogere algemene serviceniveaus. De toegevoegde waarde van het gebruik van de winkel of het distributiecentrum voor de afhandeling van online bestellingen hangt af van het verschil in de afhandelingskosten tussen de locaties.

In dit proefschrift werden MDPen gebruikt voor het oplossen van de voorraadbeheerbeslissingen van de winkelier. Aangezien het oplossen van de MDP resulteert in het optimale beleid, kan het worden gebruikt als maatstaf voor andere methoden om de prestaties te vergelijken. Het voordeel van een MDP is dat het rekening kan houden met zowel huidige als toekomstige uitkomsten, waardoor het beter presteert dan eenvoudiger methoden. Wij vergeleken onze methoden uitvoerig met die in de praktijk of in de literatuur zijn gevonden, en zijn in staat hogere winsten en service niveaus te bereiken. Bovendien is het voordeel van MDP dat men de structuur van het optimale beleid kan identificeren en er mogelijk eenvoudige, maar effectieve regels uit kan afleiden.

Het gepresenteerde onderzoek in dit proefschrift heeft het potentieel aangetoond van het gebruik van de winkel voor de activiteiten van een omni-channel retailer. Hoewel in de literatuur vaak wordt besproken over dat het gebruik van winkelvoorraad voor het online kanaal de winkelervaring van de winkelklant negatief zou kunnen beïnvloeden, heeft dit proefschrift aangetoond dat met een goede voorraadbeheerstrategie de prestaties van beide kanalen kunnen worden verbeterd. Wanneer de winkelier over meerdere winkels beschikt om aan de vraag van het online kanaal te voldoen, verhoogt dit bovendien de flexibiliteit van de vervulbeslissing. Dankzij deze grotere flexibiliteit kan de winkelier beslissen waar hij zijn voorraad plaatst, aangezien er meerdere locaties kunnen worden gebruikt voor het online vraagkanaal. Niet alleen kan de winkelier hierdoor overtollige voorraad op bepaalde locaties verminderen, de winkelier kan ook locaties met lagere voorraden beschermen door de vraag te sturen naar andere locaties.

Publications by the Author

In this dissertation

Goedhart, J., Haijema, R., & Akkerman, R. (2022). Inventory rationing and replenishment for an omni-channel retailer. *Computers & Operations Research*, *140*, 105647

Goedhart, J., Haijema, R. & Akkerman, R., (2023). Modelling the influence of returns for an omni-channel retailer. *European Journal of Operation Research*, *306*, 1248-1263

Goedhart, J., Haijema, R., Akkerman, R., & S. de Leeuw (*Under review*). Replenishment and fulfilment decisions for stores in an omni-channel retail network.

Goedhart, J., Haijema, R., Akkerman, R., & S. de Leeuw (*Under review*). Inventory and fulfilment decisions for different omni-channel concepts in a retail network setting.

Other scientific publications

Goedhart, J., & Hendrix, E. M. (2022). On computational procedures for optimising an omni-channel inventory control model. *Computational Science and Its Applications – ICCSA 2022*, Springer.

Beames, A., **Goedhart, J.**, & Kanellopoulos, A. (2021). Biobased economy: Critical foundation for achieving sustainable development goals. *Decent Work and Economic Growth*, Springer.

Dankwoord

Ik eindig mijn proefschrift met het belangrijkste en meest gelezen stuk tekst. Ik wil beginnen met René en Renzo, bedankt voor de steun en begeleiding gedurende mijn promotietraject. Onze meetings waren naast nuttig ook altijd erg gezellig en ik kan jullie na vier jaar naast begeleiders ook vrienden noemen. René, bedankt dat je me altijd methodologisch stimuleerde om zo buiten mijn comfortzone nieuwe methodes op te pakken. Ook bedankt dat ik zowel werk als privézaken met je kon bespreken. Renzo, bedankt voor alle steun en voornamelijk het verbeteren van mijn Engels. Ik kon altijd op je vertrouwen en je was bereid om elk moment van de dag me te helpen.

Verder wil ik Sander bedanken voor het meedenken tijdens mijn promotie. Toen we elkaar in Mallorca hadden ontmoet had ik al interesse om met je te samenwerken omdat ik onder de indruk was van je kennis. Dat je uiteindelijk mijn promotor zou worden ging boven mijn verwachtingen. Ik wil ook Jacqueline bedanken, die helaas niet meer bij ons is. Zonder Jacqueline was dit allemaal nooit mogelijk geweest, haar acquisitie bij het digital society initiatief waaruit dit project is gefinancierd zal ik haar voor altijd dankbaar zijn. Bedankt dat je mij dit project toe vertrouwde, het is gelukt.

Ik wil graag alle PhD'ers bedanken die ik tijdens mijn promotietraject heb ontmoet, jullie hebben mij het meest geholpen om het werk leuk te houden. Ik wil daarvoor ook meteen excuses aanbieden aan alle andere collega's in de gang voor de geluidsoverlast van gelach en geklets uit 6029. Ik wil met name enkele mensen benoemen, Joep en Maarten, ik vond het fijn dat we elkaar hadden om over alles rondom ons werk te kletsen, al ging het uiteindelijk vaak over bier. Verder wil ik Heleen bedanken, ondanks dat we niet lang kantoorgenoten zijn geweest zie ik de korte periode als een periode met veel gelach en gegiechel. Als laatste wil ik Maike en Matteo bedanken dat jullie altijd wouden luisteren naar mijn belachelijke verhalen en daarover konden lachen.

Verder wil ik natuurlijk alle collega's van ORL bedanken voor de gezellige lunches, uitjes, en gesprekken bij de koffiemachine. Jullie hebben mij indirect geïnspireerd en gesteund gedurende mijn promotie met alles. Helaas zal de gang zonder mij wat minder modieus zijn, dus hopelijk pakt een van jullie dat op. Claudia, Marieke en Laura, bedankt voor jullie ondersteuning en nog meer voor het geroddel, gelach en gekletst met jullie. Deze *fashion boy* vertrekt dus nu moet je complimenten over kleding ergens anders zoeken.

Aan al mijn vrienden, enorm bedankt voor de steun gedurende mijn promotie. Excuses dat jullie soms het gezeik moesten aanhoren als dingen niet liepen zoals ik wou maar jullie luisterende oren hebben me enorm geholpen. Mijn Wageningse vrienden, we zijn al meer dan tien jaar bevriend en ik zou dat graag willen kwadrateren. Dank dat jullie altijd klaar voor me stonden en dat we er altijd voor elkaar zijn. De Moatschap, vanaf het begin van de studie maakte jullie al grappen dat ik een PhD zou doen, nou hier is die dan uiteindelijk. Bedankt, voor de steun en nog meer de leuke avonden (die vaak onbedoeld ontspoorde), ik ben blij dat we na de studie nog steeds bevriend zijn. Als laatste Coen en Youri, bedankt voor de toepen en bollen avonden die ik nauwelijks won. Maar het gelach over onze verhalen maakten de verliezen meer dan waard.

Als allerlaatste wil ik eindigen met mijn familie, waarbij ik mijn zus Babette als allereerste wil bedanken voor de steun gedurende de tijd, waaronder de doordeweekse etentjes bij pap en mam. Ik vond het fijn dat ik mijn programmeer irritaties kon bespreken en dat je altijd meedacht. Verder wil ik Adrienne en Sten bedanken voor de gastvrijheid in Amsterdam waar ik altijd langs kon als ik Wageningen even wou ontsnappen. Jullie fantastische kookkunsten maakte een zure werkweek weer goed. Verder, Vita bedankt voor de afleiding met je vrolijkheid voor als ik het even lastig had.

Pap, bedankt voor het altijd steunen en motiveren gedurende mijn promotie. Je hebt me gemotiveerd om hiervoor te gaan en bij elke problemen en frustraties dacht je altijd met mee. Het is fijn dat ik altijd bij je terecht kon over van alles.

Als laatste wil ik mijn favoriete collega van de WUR bedanken, mam bedankt voor de steun en de wandelingen tijdens werk. We zijn ook nog een tijd kantoorgenoten van elkaar geweest waarbij we elkaar hebben geholpen om vanuit thuis ons werk te doen. Ik kijk met plezier terug op die bijzondere tijd.

Pap, mam, ik houd van jullie.

Completed Training and Supervision Plan

Joost Goedhart
Wageningen School of Social Sciences (WASS)

Learning activity	Department/Institute	Year	ECTS*
A) Project related competences			
A1 Managing a research project			
WASS introduction course	WASS	2019	1
Writing research proposal	WASS	2019	6
'Rationing Inventory Strategy for an Omni-channel Retailer'	International Working Seminar on Production Economics, Innsbruck, Austria	2020	1
'Inventory Rationing and Replenishment for an Omni-channel Retailer'	Best student paper finals, International Society of Inventory Research, Budapest, Hungary	2020	1
'Influence of Returns for an Omni-channel Retailer'	WASS PhD day, Wageningen, the Netherlands	2021	0.5
'Influence of Returns on an Omni-channel Retailer's Inventory Policy'	European Operational Research Societies conference, Athene Greece	2021	1
A2 Integrating research in the corresponding discipline			
Fundamental Knowledge on Transport, Infrastructure & Logistics	TRAIL	2018	0.5
Machine Learning with Python: from Linear Models to Deep Learning	MITx	2019	5
Reinforcement Learning Specialization	University of Alberta	2020	4
Machine Learning in OR	GP-OML	2020	2
Periodic Review Inventory Management and Markov Decision Processes	GP-OML	2021	0.5
B) General research related competences			
B1 Placing research in a broader scientific context			
Analyze data with R	Codecadamy	2021	2.9
Learn C++	Codecadamy	2021	1.2
Computational social science methods	University of California	2022	3

Continued on next page

Learning activity	Department/Institute	Year	ECTS*
B2 Placing research in a societal context			
Writing blogs	WUR	2022	1
C) Career related competences			
C1 Employing transferable skills in different domains/careers			
Teaching assistance Decision Science 2 and Operation Research and Logistics	WUR	2019-2022	2
Supervision of thesis students	WUR	2019-2022	2
Total			34.6

*One credit according to ECTS is on average equivalent to 28 hours of study load

This thesis is part of the PhD thesis series of the Beta Research School for Operations Management and Logistics (onderzoeksschool-beta.nl) in which the following universities cooperate: Eindhoven University of Technology, Ghent University, Maastricht University, Tilburg University, University of Twente, VU Amsterdam, Wageningen University & Research, KU Leuven, Universiteit Hasselt, VU Brussels, University of Antwerp and CWI, the Dutch national research institute for mathematics and computer science.

Printed by ProefschriftMaken

Cover design by Lucas van Hapert