Three Essays on Time Window Assortment Design for Grocery Home Delivery

Inaugural Dissertation
to Obtain the Academic Degree of
a Doctor in Business Administration
at the University of Mannheim

submitted by

Katrin Waßmuth, M.Sc.
Mannheim

Dean: *Joachim Lutz*
Referent: *Prof. Dr. Moritz Fleischmann*
Co-Referent: *Prof. Dr. Cornelia Schön*

Day of oral examination: 12 December 2025 to my parents, Ute and Berthold

Acknowledgements

This dissertation would not have been possible without the inspiration, encouragement, and support of many people who have accompanied me over the past five years. I am deeply grateful to all those who have contributed to this journey, whether through academic guidance, personal support, or shared moments that made this time both meaningful and memorable.

First of all, I would like to thank my doctoral advisors, Moritz Fleischmann and Niels Agatz. I am deeply grateful for their continuous support and guidance throughout this journey. I am especially thankful to Moritz for his help in identifying relevant and meaningful research questions. His thoughtful advice helped me decide on a research path to take and gave direction to my work. I particularly appreciated the freedom he granted me throughout the research process, as well as the trust-based work environment he created, one that encouraged independence while always offering insight when needed. I am also grateful to Niels for taking on the role of my co-supervisor throughout my entire doctoral journey. Thank you for the opportunity to spend time as a visiting researcher at the Rotterdam School of Management and for your encouragement and support during my first conferences. I especially appreciated our "thinking-out-loud" research meetings, which showed me the value of open, unfiltered discussions. These conversations not only shaped the quality of our work but also taught me a great deal about academic collaboration and helped me grow into a better supervisor myself. I would like to thank Cornelia Schön for co-refereeing this dissertation and Claudius Steinhardt for serving on my doctoral defense committee.

I am thankful for the academic community and the research friendships that have made this journey so much more enjoyable. I would like to thank Charlotte Köhler for the supportive and cheerful friendship that grew out of our joint project. Our meetings helped me navigate the early phases of my doctoral studies and were especially valuable during the challenges of the pandemic, when I could always look forward to our regular online meetings. Even more special was the moment we first met in person, marking the beginning of our friendship. I also want to thank Liana van der Hagen, not only for being a wonderful academic sibling but also for being the best office mate I could have wished for during my stay at RSM. I am grateful for all the fun times we have had at conferences, workshops, and outings, and for the special moments we have shared in life, hopefully with many more to come! Thanks also to the entire TOM department at RSM for giving me such a warm and welcoming experience during my time in Rotterdam, and to the last-mile community for being such supportive and inspiring academic peers.

I am also thankful to the many colleagues, peers, and friends at the chair and across the faculty who made my time in Mannheim both professionally enriching and personally enjoyable. I want to thank Müge Çakan, not only for being an amazing office mate but also for the friendship that grew out of it, and Hossein Nikpayam for his continued friendship and support. Thanks to Tobias Breitenbach, Mohammadjalal Mirbeygi-Shahabad, Pardis Sahraei, Jonas Schwamberger, and Jan Strackbein, as well as our secretaries Julia Peemöller and Ruth Pfitzmann, for being such great colleagues. I would also like to thank Borislav Bjelicic for the pleasant time working together and for his continuous support and mentorship, which I greatly appreciated. I also appreciate the shared moments with Amir Foroughi, Siamak Khayyati, Marius Krömer, Santanu Kundu, Ömer Schmiel, Niloufar Sadeghi, and Ruth Schültken. It was a pleasure having you as peers in Mannheim.

I could not have completed this journey without the enjoyable moments outside of work shared with my longtime friends. Johannes Dippert, Anna-Lena Haug, Viola Höger, Laura Penzel, Isabelle Puccini, and Ricarda Vogg: I am extremely grateful for the unwavering support I have always received from you, despite the geographical distances, as well as for the many amazing moments, conversations, activities, vacations, and life chapters we have shared and will continue to share together. I would also like to thank Lisa Fuhrmann, Annalena Mickel, and Jannik Neumann for our continued friendship since our Bachelor days, which I deeply value.

Finally, I want to express my deepest appreciation to my family. My parents, Ute and Berthold, for their unconditional love and support, for enabling me to pursue my goals, and for shaping the person I am today. My sister, Tamara, for exploring new paths alongside me, sharing many experiences, even if sometimes apart, and always having my back. My cousins, Franziska and Nadine, for continuing to value and strengthen our family bonds, and my aunts and uncles Anne, Armin, Brigitte, and Johannes for all the uplifting and meaningful conversations and for always welcoming me with open arms whenever I return home. Finally, I want to thank Florian for making the process of finishing this thesis so much more enjoyable and for helping me navigate what comes next. Thank you for always being there for me and for the memorable moments we have shared and will share together.

Summary

In online grocery retailing, customers are usually offered a set of delivery time windows to choose from, a last-mile delivery concept commonly referred to as *attended home delivery*. From a service perspective, customers tend to prefer short time windows. Retailers, however, must fulfill orders efficiently despite the inherent complexity of planning time-constrained deliveries, which rather favors longer time windows. The design of delivery time windows therefore plays a crucial role in shaping both customer demand and operational efficiency. The growing body of literature on *demand management* in attended home delivery addresses this trade-off by optimizing the offering and pricing of delivery time windows. However, existing research has primarily focused on selecting an efficient subset of time windows from a given set within medium- and short-term planning horizons. The strategic design of these sets, referred to as *time window assortments*, remains largely underexplored. This dissertation addresses a novel planning problem: how to design time window assortments that account for customer preferences while maintaining delivery efficiency. Across three essays, we address this research gap, propose a modeling approach to analyze the problem, derive tractable analytical results and obtain numerical insights to examine key trade-offs, and discuss practical implications. The first essay provides a structured literature review of demand management in attended home delivery, covering strategic, tactical, and operational approaches such as pricing, availability control, and feasibility assessment. We synthesize insights across fields of application, highlight common modeling choices, and identify promising research gaps. The second essay investigates profit implications of different time window assortments. We develop an evaluation model based on continuous approximation to assess metrics like delivery cost and capacity and examine how the number, length, and overlap of time windows affect performance. The results offer guidance for aligning operations strategy with relevant market conditions to make grocery delivery services economically viable. The third essay incorporates stochasticity in customer choice behavior across the time window assortment to further analyze the effects of overlapping time windows. We derive optimal demand allocations, identify ex-post conditions under which overlapping time windows reduce delivery costs, and apply Monte Carlo estimation to test these conditions at the decision-making level. Together, these essays provide a novel perspective on time window assortment design as a strategic lever in attended home delivery. Our findings offer theoretical insights and actionable guidance for online grocery retailers seeking to align customer satisfaction with efficient last-mile operations.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADV	Advanced routing methods
AHD	Attended home delivery
ANS	Adaptive neighborhood search
APR	Functional approximation
B2B	Business-to-business
CA	Continuous approximation
ESMR	Expected marginal seat revenue
EXO	Exogenous substitution rates
GAM	Generalized attraction model
INS	Insertion heuristics
MIP	Mixed-integer programming
ML	Machine learning
MNL	Multinomial logit
ROUTE	Explicit routing
RUT	Random utility theory
SEED	Seed-based routing
SIM	Simulation
TSP	Traveling salesperson problem
VFA	Value function approximation
VRP	Vehicle routing problem

Chapter I

Introduction

Grocery home delivery, commonly referred to as *e-grocery*, is a business model in which customers order groceries online and have them delivered directly to their homes. Although e-grocery has become widespread and customers are increasingly adopting it as a service, the sector continues to face several persistent challenges. First, the sector is marked by intense competition, not only among e-grocery providers but also with established brick-and-mortar supermarkets. Second, as a consequence of this competitive pressure, product margins in the grocery sector are typically low, leaving little room for inefficiencies. Third, e-grocery entails particularly complex last-mile logistics. Grocery items often require careful handling and must therefore be handed over to the customer in person, a requirement known as *attended home delivery*.

The challenge of balancing an attractive service offering with cost-efficient operations in e-grocery has a long history, beginning in the mid-1990s when early pioneers attempted to digitize the grocery shopping experience. One of the first movers was *Peapod*, founded in 1989, which initially partnered with traditional supermarkets and fulfilled customer orders from in-store inventory. Another early entrant was *NetGrocer*, launched in 1995, which adopted a mail-order model based on a central warehouse and shipped non-perishable groceries to customers via standard parcel carriers such as *FedEx*. The most ambitious early player was *Webvan*, founded in 1996. *Webvan* aimed to build a fully vertically integrated operation with state-of-the-art automated warehouses and its own fleet of delivery trucks, promising deliveries within narrow 30-minute time windows, which is an impressive target even by today's standards. The company raised nearly $400 million in venture capital and went public in 1999 with a valuation of over $4 billion. However, *Webvan* became one of the most prominent failures of the dot-com era. Its rapid expansion into multiple cities, ahead of establishing a proven, profitable model, led to massive overhead and underutilized fulfillment centers. Demand failed to meet expectations, and the company was burning through cash far faster than it was generating revenue. In 2001, just two years after its initial public offering, *Webvan* filed for bankruptcy.

From today's perspective, *Webvan* can be seen as a business that was ahead of its time, with a model that proved too ambitious given consumer behavior and technological infrastructure in the late 1990s. Its failure served as a cautionary tale, highlighting the risks of scaling too quickly without first establishing operational viability. Two decades later, a similar pattern emerged with the rise of quick-commerce startups such as *Gorillas* and *Flink*, which promised ultra-fast delivery within 10 to 15 minutes. *Gorillas* expanded rapidly across Europe while burning through investor capital at unsustainable rates, eventually withdrawing from multiple markets before filing for insolvency in 2023. In contrast, *Flink* remains operational and appears to be improving efficiency by focusing on better order consolidation and route planning, accepting longer delivery time windows in exchange, suggesting a more sustainable path in the competitive quick-commerce sector. Similarly, another segment of e-grocery providers prioritizes operational robustness, route optimization, and higher customer densities to achieve long-term profitability. Companies such as *Instacart* and *Amazon Fresh* have partnered with existing retailers or leveraged freelance shoppers, enabling more sustainable scaling. Other successful models include *Picnic* and *Albert Heijn*, both of which use optimized time window management and routing. *Picnic* is a pure-play e-grocer operating through centralized fulfillment, while *Albert Heijn* leverages its existing retail network by integrating physical stores with dedicated fulfillment centers to efficiently fulfill online orders.

Effectively managing attended home deliveries requires service providers to make a series of interdependent decisions, one of which concerns the timing of deliveries. To ensure a successful handover, the service provider and the customer must agree on a suitable delivery time. In many established business models, this agreement is facilitated through a selection of delivery time windows offered to customers, with each time window corresponding to an upcoming delivery shift. Once customers have made their choices and the order cut-off time has passed, the provider plans the corresponding delivery shift. This involves solving a *Vehicle Routing Problem with Time Windows*, a computationally demanding yet well-established optimization task that schedules deliveries efficiently while respecting the selected time windows. Customer preferences for delivery time windows and the provider's efficiency considerations give rise to a fundamental cost-service trade-off: Short windows enhance customer satisfaction, whereas longer windows facilitate more efficient route planning.

Consequently, *time window management* has emerged as a key and extensively studied lever in the literature on attended home delivery. Due to its ability to shape demand, it is often considered a form of revenue management, while customers' time window choices also influence the provider's ability to plan efficient delivery routes. The challenge lies in balancing these competing objectives: maximizing customer appeal without compromising logistical efficiency, which is further complicated by uncertainty in customer behavior and the structural complexity of delivery operations. Existing literature on time window management focuses almost exclusively on operational and tactical approaches, such as limiting the options available to customers or adjusting delivery fees, for a given set of time windows. However, the impact of the strategic choice of this set, referred to as the *time window assortment*, remains largely underexplored. This dissertation addresses this strategic decision by posing the following research question: How do the characteristics of a time window assortment, including its width, the length of its time windows, and whether the time windows in the assortment overlap, affect the service provider's profit and operational efficiency?

In the first essay, presented in Chapter II, we provide a comprehensive classification and review of research on time window management in attended home delivery. We begin by categorizing existing business models based on their delivery strategy. Next, we map the inherent trade-offs between satisfying customer preferences and maintaining operational profitability across different planning levels and demand management levers. Our review shows that operational demand management is the most extensively studied area, with research primarily focusing on sophisticated solution methods for selected parts of the real-time decision problem. In contrast, contributions on tactical demand management are less frequent, yet they address a broader range of decisions, from long-term customer agreements to short-term availability control. Contributions to strategic demand management appear to be sparse and diverse. Among the identified future research directions, strategic demand management offers a particularly promising opportunity to provide decision support that maximizes demand potential and ensures profitability, serving as the central motivation for the research question addressed in this dissertation.

In the second essay, presented in Chapter III, we introduce a novel *time window assortment evaluation model* that captures demand-side responses while approximating delivery costs through continuous approximation. Using this model, we assess how the number, length, and potential overlap of time windows affect demand and delivery efficiency under homogeneous conditions and stylized demand patterns, deriving tractable expressions that rely on only a few key parameters. Our analytical results identify which time window assortments can achieve profitability under different market conditions, highlighting minimum time window lengths, break-even demand thresholds, and scenarios in which maximum vehicle capacities or tour utilization become limiting factors. By endogenizing demand, we find that for customers who are insensitive to the assortment, providers should offer the longest time windows that still allow efficient tour utilization, and the choice of the number of consecutive time windows should balance the trade-off between demand clustering and spreading, while introducing overlapping time windows reduces profit. For assortment-sensitive customers, an optimal time window length balances demand attraction with the required tour frequency, and adding additional time windows, whether consecutive or overlapping, can increase profit if the incremental demand is sufficient.

In the third essay, presented in Chapter IV, we extend the analysis of time window assortment design by examining overlapping time windows, a common but underexplored structure, under demand uncertainty. We focus on their potential to mitigate variability in demand distribution across time windows. Demand is modeled as a random variable capturing both variability and differences in time window popularity, and we derive conditions under which overlaps reduce delivery costs compared to consecutive designs. These results inform hypotheses that we test using a Monte Carlo simulation to quantify the expected performance gap. Our findings show that overlaps can smooth operational bottlenecks by reallocating workload across adjacent time windows, but their effectiveness depends on operational parameters, demand characteristics, and customer behavior. They are most beneficial when service time dominates routing time, total demand is high, and demand variability is significant. Their success also hinges on the interaction between demand transition rates and initial demand distributions, and in some cases, overlaps may even create new bottlenecks. Overall, overlapping windows should be treated as a strategic design choice whose value depends on expected demand patterns and customer preferences.

This research suggests several promising future directions. Incorporating more sophisticated models of customer behavior, including endogenous choice and transition dynamics, would enable a nuanced assessment of time window assortments, capturing both cost and revenue implications. Further research should explore the interaction between strategic time window assortment decisions and tactical or operational levers as a means to smooth demand over time and align offerings with customer preferences, thereby improving operational efficiency and revenue potential. Another avenue is to extend the evaluation of time window assortments beyond profitability by incorporating environmental and social sustainability metrics. Finally, empirical validation using operational data is crucial to test model assumptions, quantify practical benefits, and provide actionable managerial guidance for grocery home delivery.

Chapter II

Demand Management for Attended Home Delivery: A Literature Review[1]

with Charlotte Köhler, Niels Agatz, and Moritz Fleischmann

Abstract

Given the continuing e-commerce boom, the design of efficient and effective home delivery services is increasingly relevant. From a logistics perspective, attended home delivery, which requires the customer to be present when the purchased goods are delivered, is particularly challenging. To facilitate the delivery, the service provider and the customer typically agree on a specific time window for service. In designing the service offering, service providers face complex trade-offs between customer preferences and profitable service execution. In this paper, we map these trade-offs to different planning levels and demand management levers, and structure and synthesize corresponding literature according to different demand management decisions. Finally, we highlight research gaps and future research directions and discuss the linkage of the different planning levels.

[1]The contents of this chapter have been published in Waßmuth et al. (2023).

2.1. Introduction

The COVID-19 pandemic has boosted the demand for online shopping and home delivery across the globe, and it is likely that some shifts in demand will also have long-lasting effects (OECD, 2020). For example, the global online share of grocery annual sales increased from 7% before the pandemic to 10% at its peak and remains at a high level of 9%, even after the peak[2]. Fulfilling this growing demand requires effective and cost-efficient last-mile delivery operations. While the last mile is generally recognized as the most challenging part of the fulfillment process, this is especially true for *attended home delivery* (AHD), where the customer must be present to receive the goods.

AHD is common for home services and products that require special handling, such as groceries, large appliances, or furniture. To reduce missed deliveries and waiting times, service providers typically let customers choose a delivery time from a menu of time windows or deadlines (referred to as service options). This step involves the customer directly in the service creation process, a characteristic that is typical of the field of service operations management (see, e.g., Coltman & Devinney, 2013).

The concept of AHD is especially well established in the context of online grocery retailing, which is a particularly challenging sector, as profit margins are low, and the delivery of fresh or even frozen goods requires special care in planning and execution. Consequently, many online supermarkets are struggling to create a profitable business[3,4]. To manage profitability, service providers can manage both supply and demand. The supply-side levers involve traditional supply chain planning tasks, such as network design, inventory management, and vehicle routing. In general, these levers seek the most cost-efficient fulfillment of a given demand (see, e.g., Han et al., 2017).

Demand management focuses on managing customer demand to maximize profitability of a given supply. Typical levers include the specific service options and prices offered to customers. Through these levers, demand management can enhance profits in two ways. First, by increasing revenues by prioritizing high-value customers or by serving more customers due to better capacity utilization. Second, demand management may reduce costs by facilitating more efficient order delivery. In addition to profit maximization, demand management can also contribute to other goals, such as prioritizing

[2]Statista, https://bit.ly/3h4kiXG. Accessed on February 14, 2022
[3]Tagesspiegel, https://bit.ly/3vpokhZ. Accessed on February 14, 2022
[4]Chicago Tribune, https://bit.ly/3t3ZXEM. Accessed on February 14, 2022

specific customer groups when demand exceeds capacity (Schwamberger et al., 2023) or steering customers toward more sustainable delivery times (N. Agatz et al., 2021).

While traditional supply-oriented approaches have been studied for decades, demand management has only started to attract substantial attention in the research community more recently. Technological advances have been driving this development by allowing for a better understanding of customer behavior and by providing the flexibility to change offered services and prices in real time. When considering current practice, we observe that different e-grocers make different choices regarding their service offerings. In the Netherlands, for example, Albert Heijn offers up to 15 different time windows per day with various lengths (one to six hours) and different delivery fees, whereas Picnic offers any customer a single, free, one-hour time window for each day of the week. We also observe a dynamic development in terms of business models, including on-demand grocery delivery, as offered by Gorillas and Flink. Given the recent progress in the field, the time appears right for a review of demand management for AHD to synthesize the current knowledge and identify relevant open questions.

Demand management generalizes the concept of revenue management, which aims to maximize revenues (A. K. Strauss et al., 2018). Costs are generally sunk or proportional to demand in traditional revenue management settings (R. Klein et al., 2020). In contrast, delivery costs in AHD cannot simply be attributed to individual orders but depend on the specific set of accepted orders (Snoeck et al., 2020). Demand management in AHD involves deciding on the *assortment* of the delivery service options. This links the topic to the field of assortment planning of physical products across different retail channels (see, e.g., Bernstein et al., 2019).

This paper contributes to the existing literature in the following ways. First, we refine and extend the framework by N. Agatz et al. (2013) and classify different demand management decisions along strategic, tactical, and operational planning levels. Thereby, our work is the first to explicate the different interrelated planning levels in demand management for AHD. Second, we structure and synthesize the current literature according to the different demand management decisions and planning levels. This provides an up-to-date overview of the literature and identifies research gaps and directions for future research. Third, we introduce a consistent terminology to help bring together different strands of research within the fields of revenue management and vehicle routing. In this way, our work complements previous review papers on online order fulfillment and customer behavior (Nguyen et al., 2018) and integrated demand and revenue management in vehicle routing (Fleckenstein et al., 2023; Snoeck et al., 2020).

The remainder of this paper is organized as follows. In Section 2.2, we define and structure the field of demand management and develop our classification framework to structure the academic research field systematically. Based on this framework, in Sections 2.3 to 2.5, we review the demand management literature in detail and cluster them into different research streams. In Section 2.6, we highlight our observations and identify gaps and future research opportunities for each planning level. We also discuss the connection between planning levels in that section. Finally, we conclude this literature review in Section 2.7 by summarizing our main findings and pointing out general avenues for future research.

2.2. Demand Management Framework

In this section, we structure the field of demand management for AHD and embed it into a planning framework. To this end, we first highlight important structural elements of the fulfillment process (Section 2.2.1). Second, we characterize the different planning levels and identify the related demand management levers (Section 2.2.2). We use the resulting framework to structure our literature review in Sections 2.3 to 2.5.

2.2.1. Order Fulfillment Process

Demand management for AHD aims to generate customer demand and, at the same time, shape it in a way that benefits the fulfillment process. To identify the potential of demand management in this context, we thus need to understand the fulfillment process. At a broad level, it involves activities in sourcing, warehousing, delivery, and sales (N. A. H. Agatz et al., 2008). However, in our context, the most relevant part of the fulfillment process is the one that follows the interaction with the customer, i.e., the customer order decoupling point. This downstream part comprises three main steps, namely, order capture, order assembly, and order delivery (Campbell & Savelsbergh, 2005). In what follows, we briefly discuss each of these steps and how to coordinate them for multiple orders.

Fulfillment Steps

During *order capture*, the customer and the service provider mutually agree on when and where the order is to be delivered. To reach such an agreement, the service provider commonly presents an assortment of service options from which the customer can choose. The offered service options may differ in their timing within and across days, their lengths, and their associated delivery prices. Some providers offer the same set of options to all customers, while others tailor them to the customer's shopping history, delivery location, or basket composition. To ensure a smooth booking process, the service provider must decide on the offered service assortment very quickly, within, at most, a few seconds. Customers choose from the offered options according to their preferences – not placing an order if none of the options meets their expectations. Once the customer chooses a service option, the service provider confirms the order, and the delivery agreement is fixed. It is illustrative to position this process relative to adjacent research fields: In the terminology of the production planning literature, the described process is denoted as real-time single-order capture (Meyr, 2009), while service operations management classifies it as nonsequential offering (Liu et al., 2019).

Order assembly denotes all warehousing operations that are required to prepare an order for delivery, including order picking, sorting, and packaging. Handling the items may be demanding depending on the product category. For example, grocery orders may contain dry, fresh, refrigerated, and even frozen food. This makes order picking quite time consuming. Many service providers therefore seek economies of scale by consolidating the order assembly in larger fulfillment centers that allow for (semi-)automated picking processes. This, however, usually moves the order assembly location further away from the delivery areas, thereby increasing the overall fulfillment lead time. Constraints on innercity space further exacerbate this effect. Service providers that compete on short click-to-door times may therefore opt for a different approach, relying on smaller fulfillment centers situated near customer locations. In particular, on-demand service providers often use a dense network of small innercity depots or even assemble orders in physical stores.

Order delivery refers to the physical delivery of purchased products to customers' homes within a certain time frame. As this step typically involves assigning customer orders to vehicles and determining the delivery sequence, it can be modeled as a vehicle routing problem (VRP). Service providers often run a proprietary delivery fleet; only a few use external carriers. The fleet can be composed of trucks, vans, cars, or bicycles that visit one or more customers along a specified route. The service includes delivery to the customer's doorstep, and thus, delivery includes a service time for handover, parking, unloading and – for apartment buildings – carrying the order upstairs. For online supermarkets, the service time is approximately 10 minutes (R. Klein et al., 2019).

Fulfillment Process Design

For a single customer order, the three steps of the fulfillment process naturally follow the sequence outlined above. However, the service provider has multiple options to coordinate these steps across multiple orders. For example, the order assembly literature discusses wave and waveless release times, where the former means that incoming orders are held back to be later released in larger batches, whereas in the latter, arriving orders are released immediately and individually (see, e.g., Çeven & Gue, 2017). Similar options apply to order delivery, as discussed in the literature on dynamic consolidation by means of dispatch waves (see, e.g., Klapp et al., 2018). For AHD, we distinguish between a periodic and order-based design of the fulfillment process.

In a *periodic* fulfillment process, the service provider defines periodic cut-off times, after which all captured orders are assembled and delivered. In other words, there is a fixed period for assembly and delivery that does not overlap with the respective order capture period. This approach exploits economies of scale by consolidating orders in the assembly and delivery steps. The resulting efficiency benefit comes at the expense of a longer click-to-door time since captured orders have to wait until the cut-off time before being further processed. The service provider can choose the cut-off frequency to manage the speed/efficiency trade-off. For online groceries, daily or semi-diurnal cut-offs are common.

In an *order-based* fulfillment process, the service provider decides dynamically on each customer request whether to initiate the assembly and delivery of orders captured up to that time. In particular, this includes the option to assemble and deliver each order individually immediately after capture. Intuitively, this process design is common for businesses that compete aggressively on speed. It is worth pointing out that a 'same-day delivery' service does not necessarily imply an order-based fulfillment process. In fact, under periodic fulfillment, a cut-off time early in the day may also allow for deliveries later on that same day. Thus, from a planning perspective, there is a greater distinction between periodic and order-based processes than between 'same-day' and 'next-day' delivery. We illustrate this point with specific examples below and visualize it in Figure 2.1.

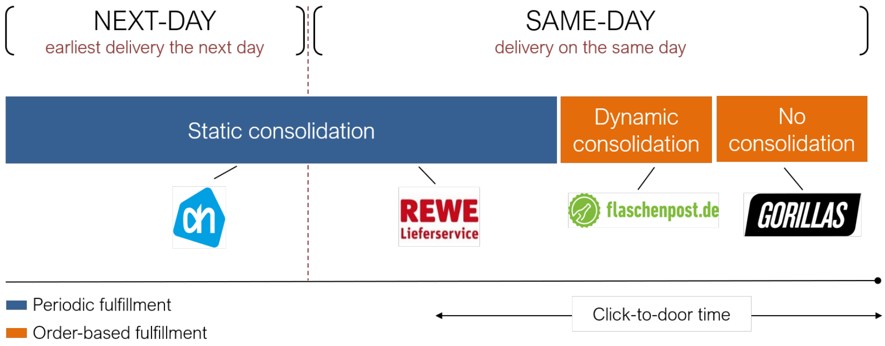


Figure 2.1.: Illustration of fulfillment process design alternatives

The Dutch grocery retailer Albert Heijn follows a periodic fulfillment process with cut-off times at noon for deliveries the next morning, and at midnight for deliveries the next afternoon[5]. After each cut-off, delivery routes are planned, and order assembly takes place in one of five online fulfillment centers[6]. Similar to Albert Heijn, the German e-grocer REWE also operates a periodic fulfillment process. REWE uses a cut-off time of 1 pm, which allows orders to be delivered in the late afternoon on the same day. To enable fast delivery and handling of more than 20,000 products, the company invests in semi-automated fulfillment centers close to delivery areas[7].

In contrast, the German beverage delivery service Flaschenpost does not communicate periodic cut-off times but guarantees delivery within 120 minutes for every incoming order – a service proposition that requires a particularly fast fulfillment process. To meet this requirement, Flaschenpost operates 23 fulfillment centers to distribute an assortment of approximately 2,000 products to more than 150 German cities[8]. Each of these facilities is equipped with approximately 70 vans that deliver up to ten orders per trip[9]. We denote this fulfillment approach as order-based with dynamic order consolidation.

[5]Albert Heijn, https://bit.ly/3gsLv6x. Accessed on February 14, 2022
[6]Ahold Delhaize, https://bit.ly/3q49Jap. Accessed on February 14, 2022
[7]REWE, https://bit.ly/2SepFd2. Accessed on February 14, 2022
[8]Flaschenpost, https://bit.ly/3gx0B9U. Accessed on February 14, 2022
[9]Flaschenpost, https://bit.ly/3xbrira. Accessed on February 14, 2022

Further speeding up the fulfillment process, German start-up Gorillas offers on-demand grocery delivery within 10 minutes. To meet the extremely short delivery times, the company sets up micro fulfillment centers in each delivery area and limits the offered product assortment to 2,500 products. In addition, they hand-pick each captured order immediately and deliver it by bicycle[10]. Such a fulfillment process is order-based without consolidation.

2.2.2. Demand Management Decisions

In the previous subsection, we highlighted the main steps of the fulfillment process in AHD services. How efficiently a company can execute these steps depends on the properties of individual orders, such as their click-to-door time (e.g., M. Ulmer, 2017) and delivery time specificity (e.g., Lin & Mahmassani, 2002), as well as on the temporal and geographical distribution of the overall set of captured orders (e.g., Ehmke & Campbell, 2014). At the same time, these factors are intimately linked to customer preferences and thus to the popularity of delivery service options. Demand management aims to manage the resulting trade-offs between captured demand (revenue) and assembly and delivery efficiency (costs). In this sense, Figure 2.2 illustrates the interdependence between demand management and the steps of the fulfillment process and the implied impact on revenue and costs.

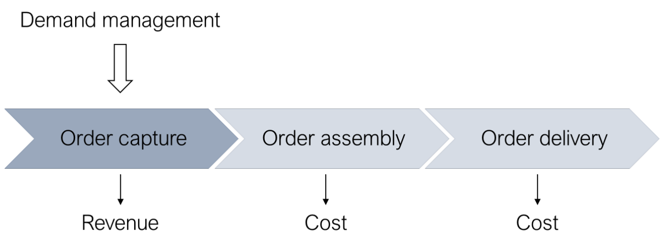


Figure 2.2.: The role of demand management within the fulfillment process

Demand management encompasses a diverse set of different decisions. We propose mapping these out along two dimensions, distinguishing three *planning levels* (strategic, tactical, and operational) and two *levers* (offering and pricing). This approach gives rise to six different sets of demand management decisions, as shown in Table 2.1. In what follows, we briefly discuss both dimensions of this framework.

[10]Supermarktblog, https://bit.ly/3eNUBsb. Accessed on February 14, 2022

Table 2.1.: Demand management framework

	Offering *design & availability*	Pricing *incentives*
Strategic *demand potential*	**Strategic offering**	**Strategic pricing**
Tactical *demand forecast*	**Tactical offering**	**Tactical pricing**
Operational *actual demand*	**Operational offering**	**Operational pricing**

Planning Levels

As is common in many areas of supply chain planning and operations management (Fleischmann et al., 2015), we distinguish between different hierarchically-linked planning levels, i.e., strategic, tactical, and operational. We define these levels based on their aim, their time horizon, and their relation to the fulfillment process timeline. Strategic decisions are design choices specified over a long horizon, while tactical and operational decisions consider the management of service options over a shorter time span. Strategic and tactical decisions take place before order capture while operational decisions are based on real-time information on actual demand. In what follows we elaborate on each of these levels in some more detail.

Strategic demand management defines the boundaries within which tactical and operational demand management are embedded. It constitutes a special case of the service design stage in service operations management (see, e.g., Roth & Menor, 2003) and also bears resemblance with structural decisions in revenue management (K. T. Talluri & Van Ryzin, 2004). Strategic demand management reflects the overall business strategy and, to gain a competitive advantage, must be carefully aligned with the competitive environment, customer preferences and willingness to pay, and operational implications. Respective decisions determine the target markets and design the general service assortment, based on a market's demand potential. This includes selecting the service region and pricing model, designing the service options, and defining appropriate service segments for subsequent tactical planning. The term service segment refers to a customer group that should receive the same service assortment (e.g., a geographical area).

The subsequent planning levels address the management of the designed service assortment within the established boundaries. We classify any such decisions taken before order capture as *tactical* demand management. Tactical decision-making is based on (aggregated) demand forecasts and exploits the heterogeneity of customers in the delivery market. Corresponding decisions include differentiation of service options and prices for different service segments. Moreover, tactical planning can be applied to simplify short-term operational planning, for which only limited computational time is available.

We denote any decisions made during order capture as *operational* demand management, i.e., decisions that are made in real time, based on detailed information on actual customer orders. Thus, operational decisions are highly time-critical and directly affect the interaction with the customer. They include accepting customer orders and adjusting the availability of service options and attached prices in the short term. For order-based fulfillment processes, these decisions are additionally combined with simultaneous fulfillment planning, as the order capture step overlaps with order assembly and delivery. This differs from periodic designs, where fulfillment planning can be postponed until after the cut-off. Both tactical and operational demand management share analogies with traditional revenue management (N. Agatz et al., 2013; Snoeck et al., 2020).

In this subsection, we introduced the planning levels top-down from strategic to operational, thereby reflecting the natural sequence of decision-making. However, we observe that the corresponding literature is evolving in the opposite direction, with many demand management approaches starting at the operational level and gradually providing insights to the strategic level. We follow this development in Sections 2.3 to 2.5 and review the demand management literature bottom-up, from operational to strategic planning.

Levers

The demand management levers, offering and pricing, capture the main characteristics of the delivery service. *Offering* refers to both the design of service options and the management of their availability. The latter are binary decisions (an option is either offered or not offered) that can (i) ensure feasibility and (ii) steer customer choice by intentionally withholding some feasible options. Service providers can also manage demand through *pricing* decisions. We use ‘pricing’ to denote a variety of monetary and non-monetary incentives to steer customer choice and generate additional revenue by exploiting differences in willingness to pay. The pricing lever allows a more fine-grained demand management since prices can be chosen from a continuous interval, rather than from a binary set. Previous research in the context of e-grocery suggests that small incentives may suffice to change customer behavior (Campbell & Savelsbergh, 2006).

Offering and pricing can be used as substitutes to steer demand. However, it should be noted that customers might perceive them very differently, as the willingness to pay is generally low (Goethals et al., 2012). Furthermore, the two levers also have complementary features and constitute building blocks that can be combined into an overarching demand management approach. For example, in the case of operational demand management, pricing usually builds on the feasibility decision, i.e., the service provider first determines which options could be offered, and then sets prices for the feasible set of options. Therefore, and in line with the dichotomy of quantity- and price-based revenue management (K. T. Talluri & Van Ryzin, 2004), we present and discuss offering and pricing separately in what follows.

2.3. Operational Demand Management

In this section, we review the literature on operational demand management, distinguishing offering and pricing decisions. We provide an overview of the corresponding literature in Table 2.2. We characterize published work with respect to the considered *problem setting*, the *decision-making process*, and the *computational study*. We further elaborate on these characteristics below. They then lead us to identifying clusters of closely related papers that we discuss in Sections 2.3.1 and 2.3.2.

We distinguish different problem settings for operational demand management by the design of the *fulfillment process* (periodic or order-based) and by the type of *service options* offered to the customer, i.e., time window or deadline.

To characterize the decision-making process, we highlight the service provider's assortment *decision approach*, that is making decisions either independently for individual service options or jointly for a set of options. Related to this aspect, some papers explicitly model customer *choice behavior*, either through exogenous substitution rates (EXO) or based on random utility theory (RUT). The remaining papers do not model customer choice but assume demand to be independent of the service offering. We also consider two attributes concerning the assessment of an incoming order. First, the service provider must verify the *fulfillment feasibility* of each service option, given the available fulfillment capacity and the previously committed orders. The feasibility check can be based on either a functional approximation (APR) or a tentative route plan, using simple insertion heuristics (INS) or more advanced routing methods (ADV). Note that for each paper only one of possibly several methods applied is given in the table. In addition to checking feasibility, the service provider may assess the present *order value* according to different metrics, including cost, service, revenue, and profit. If no order value is considered, they make decisions based on feasibility only. Papers also differ in the components of the fulfillment process that they consider in the assessment of the current order. These may include subsequent *order assembly* and *order delivery*. In addition, papers may or may not consider the impact on the fulfillment of future orders, reflected in *opportunity costs*.

For the computational study, we list the type of *demand data* (synthetic or empirical) and the *business sector* of the motivating application.

2.3.1. Operational Offering

Operational offering decisions determine the service options to offer to a customer during order capture. To structure our discussion, we cluster papers with similar characteristics as shown in the upper part of Table 2.2. In particular, we identify three clusters based on the design of the *fulfillment process* and the consideration of *opportunity costs* in the order assessment.

Periodic Fulfillment with Focus on Order Delivery Assessment Most papers that focus on operational offering decisions consider periodic fulfillment. We can further classify these papers based on whether or not they take into account opportunity costs and thus future orders. Table 2.2 shows that the papers that ignore the opportunity costs consider single time windows independently and do not explicitly model customer choice behavior. Most of these papers focus on assessing fulfillment feasibility.

One of the challenges of integrating routing aspects into operational demand management is to quickly obtain good solutions to allow for real-time feasibility checks. Hungerländer et al. (2017) develop an adaptive neighborhood search heuristic (ANS) to determine feasible time windows during order capture. The authors tailor their ANS to the specific time window problem structure to find better solutions in less time. Truden et al. (2022) study a number of different solution methods for the AHD setting. In line with Hungerländer et al. (2017), they show that it is beneficial to adapt time window

Table 2.2.: Operational demand management

		Problem setting		Decision-making process							Computational study	
				Assortment decision		Order assessment		Order assessment components				
		Fulfillment process	Service options	Decision approach	Choice behavior	Fulfillment feasibility	Order value	Order assembly	Order delivery	Opportunity costs	Demand data	Business sector
Offering	Ehmke and Campbell (2014)	Periodic	Time window	Indep.	–	INS	–		✓		Synthetic	E-grocery
	Casazza et al. (2016)	Periodic	Time window	Indep.	–	ADV	Service		✓		Synthetic	Service
	Hungerländer et al. (2017)	Periodic	Time window	Indep.	–	ADV	–		✓		Synthetic	E-grocery
	Köhler and Haferkamp (2019)	Periodic	Time window	Indep.	–	APR	–		✓		Empirical	E-grocery
	Visser et al. (2024)	Periodic	Time window	Indep.	–	ADV	–		✓		Synthetic	E-grocery
	Köhler et al. (2020)	Periodic	Time window	Indep.	–	INS	Cost		✓		Empirical	E-grocery
	Truden et al. (2022)	Periodic	Time window	Indep.	–	ADV	–		✓		Synthetic	E-grocery
	van der Hagen et al. (2022)	Periodic	Time window	Indep.	–	APR	–		✓		Synthetic	E-grocery
	Campbell and Savelsbergh (2005)	Periodic	Time window	Indep.	–	INS	Profit		✓	✓	Synthetic	E-grocery
	Mackert (2019b)	Periodic	Time window	Joint	RUT	INS	Profit		✓	✓	Synthetic	E-grocery
	Avraham and Raviv (2021)	Periodic	Time window	Joint	RUT	INS	Cost		✓	✓	Synthetic	Service
	Lang et al. (2021)	Periodic	Time window	Joint	RUT	INS	Profit		✓	✓	Empirical	E-grocery
	Lang et al. (2021)	Periodic	Time window	Joint	RUT	APR	Profit		✓	✓	Empirical	E-grocery
	Azi et al. (2012)	Order-based	Time window	Indep.	–	ADV	Profit	✓	✓	✓	Synthetic	E-grocery
	Klapp et al. (2020)	Order-based	Deadline	Indep.	–	ADV	Cost	✓	✓	✓	Synthetic	Same-day
Pricing	Campbell and Savelsbergh (2006)	Periodic	Time window	Joint	EXO	INS	Profit		✓		Synthetic	E-grocery
	Yang et al. (2016)	Periodic	Time window	Joint	RUT	INS	Profit		✓	✓	Empirical	E-grocery
	R. Klein et al. (2018)	Periodic	Time window	Joint	RUT	INS	Profit		✓	✓	Synthetic	E-grocery
	Koch and Klein (2020)	Periodic	Time window	Joint	RUT	INS	Profit		✓	✓	Synthetic	E-grocery
	Asdemir et al. (2009)	Periodic	Time window	Joint	RUT	APR	Revenue		✓	✓	–	E-grocery
	Yang and Strauss (2017)	Periodic	Time window	Joint	RUT	APR	Profit		✓	✓	Empirical	E-grocery
	Vinsensius et al. (2020)	Periodic	Time window	Joint	EXO	–	Profit		✓	✓	Synthetic	E-grocery
	Lebedev et al. (2021)	Periodic	Time window	Joint	RUT	APR	Profit		✓	✓	–	E-grocery
	A. Strauss et al. (2021)	Periodic	TW bundle	Joint	RUT	APR	Profit		✓	✓	Synthetic	E-grocery
	Prokhorchuk et al. (2019)	Order-based	Deadline	Joint	RUT	INS	Profit		✓	✓	Synthetic	Same-day
	M. W. Ulmer (2020)	Order-based	Deadline	Joint	RUT	INS	Profit		✓	✓	Synthetic	Same-day
	V. Klein and Steinhardt (2023)	Order-based	Deadline	Joint	RUT	ADV	Profit		✓	✓	Synthetic	Same-day

heuristics to the specific problem settings. Köhler and Haferkamp (2019) compare various vehicle routing methods to facilitate fast high-quality assessments of the available fulfillment capacity. The authors also introduce an acceptance mechanism based on C. F. Daganzo (1987) to approximate expected travel times. Using real-world booking data of an online supermarket, they show that the delivery area and expected demand impacts the performance of different approaches. Visser et al. (2024) study a setting in which multiple customers interact with the booking system simultaneously. It is therefore not only important to do a fast initial feasibility check but also a second check when the customer commits to a certain time window. Their detailed computational study shows that combining a fast insertion heuristic with a sophisticated background procedure ultimately leads to more accepted orders. van der Hagen et al. (2022) study the use of machine learning (ML) methods to predict the fulfillment feasibility by framing the problem as a binary classification problem. Their results suggest that ML methods can generate accurate feasibility assessments in a fraction of the time needed for common heuristic-based methods.

Another challenge of delivery-oriented order assessment is to account for uncertainty at the time of decision-making. Ehmke and Campbell (2014) seek a reliable feasibility assessment in a setting with uncertain travel times. They compare assessment methods, including a novel insertion-based heuristic that accounts for time-dependent and stochastic travel times. Based on a computational study using real travel data, they find that considering time-dependent travel times is especially valuable in suburban areas, whereas buffers against travel time uncertainty are effective in downtown areas. In addition to feasibility checks, some papers also estimate the present order value using cost and service metrics to maximize the number of orders accepted. In contrast to the cost metric, the service metric explicitly measures customer satisfaction with respect to the service options. Casazza et al. (2016) try to insert a new customer into the current route plan. If this is infeasible, the service provider does not reject the order, but shifts or enlarges the delivery time window. The authors use a dynamic programming algorithm to assess feasibility in real-time and evaluate several decision policies based on different service measures. The results highlight the trade-off between customer service and increasing the number of accepted orders. Köhler et al. (2020) introduce flexibility mechanisms that incorporate myopic information about routing efficiency and delivery locations to dynamically decide whether to offer a long or short time window to a given customer. Their results confirm that the more customers book long time windows, the more flexibility can be maintained for the fulfillment, which increases the availability of time windows for later customers.

Periodic Fulfillment with Opportunity Cost Assessment Within the second cluster, we find literature that considers opportunity costs in the assessment of a given order so as to better steer customers to more profitable or cost-efficient options. Contrary to the first cluster, most of the papers simultaneously consider multiple time windows and explicitly model customer choice behavior. However, the techniques applied to test fulfillment feasibility are simpler than in the previous cluster.

In contrast to other papers in this cluster, Campbell and Savelsbergh (2005) decide on individual time window offers independently but are the first to provide a rough estimate of future profits. In particular, for each new request, they solve a routing instance including already accepted customers, the current customer under consideration, and a number of expected future customers.

The remaining papers explicitly model customer choice behavior based on random utility theory. Incorporating customer choice behavior is crucial for joint assortment decisions. However, it is challenging to incorporate a detailed customer choice model taking into account choices and substitution across multiple days, time windows, and delivery prices. Therefore, these models try to balance modeling detail and computational effort. To this end, Mackert (2019b) apply a generalized attraction model (GAM) which ranks each time window offer based on the customer's perceived attractiveness. The authors use the choice probabilities in combination with a mixed-integer programming (MIP) based profit estimation to determine the subset of most profitable time windows for a given customer. They conclude that applying the GAM can lead to a more accurate estimation of customer choice than applying the most frequently used multinomial logit (MNL) model (e.g., Avraham & Raviv, 2021; Lang et al., 2021, 2021). Lang et al. (2021) propose several methods for anticipatory profit estimation using, inter alia, extensive offline training based on samples of expected demand and value function approximation (VFA; see, e.g., Powell, 2016). They highlight the modular composition of the associated routing and revenue management techniques. Lang et al. (2021) additionally account for multiple short- and long-term revenue metrics, including basket value, the visibility of branded trucks, and popularity among influential customers. In contrast, Avraham and Raviv (2021) focus on efficient multi-day assortment decisions. Different from the previous work, the authors anticipate future demand to maximize the number of expected accepted customers. They use tentative route information both for feasibility checks and as features of a VFA jointly predicting route efficiency over multiple consecutive days. The presented results show that taking into account inter-day dependencies create more efficient fulfillment routes that allows for more accepted orders.

Order-based Fulfillment The third cluster addresses offering decisions in order-based fulfillment systems. To date, only a few publications pertain to this stream of demand management literature. The work in this cluster presents sophisticated order delivery methods for order assessment and also takes rough proxies of order assembly into account. We conjecture that the importance of considering all fulfillment steps in the offering decision stems from the order-based fulfillment setting itself, and is due to the high time pressure in this setting.

Azi et al. (2012) consider a setting in which new customer requests arrive during the execution of the routes of previously accepted customers. There are no predetermined cut-off times. However, new customers can only be inserted into time windows of routes that have not yet started. To the best of our knowledge, this is the first paper to integrate vehicle dispatching and order capture. By assuming a load-dependent setup time, this paper also models the interaction between order capture and order assembly. The authors formulate a dynamic decision model in which the acceptance of a customer request depends on a scenario-based opportunity costs. The embedded routing problem is solved with an ANS heuristic. Instead of time windows, Klapp et al. (2020) consider the acceptance of requests that must be delivered no later than the end of the operating day, which constitutes a common delivery deadline. The objective is to minimize the sum of expected travel costs and penalties for rejecting a request. The authors approach this problem as an extension to the dynamic dispatch waves problem (Klapp et al., 2018), adding efficient request acceptance as a demand management decision. They evaluate fulfillment feasibility based on dispatch plans that include a constant parameter representing assembly time, and construct and upgrade the plans using neighborhood search heuristics.

2.3.2. Operational Pricing

Operational pricing involves dynamically adjusting the prices of the service options offered during the order capture step. This means setting (customer-specific) delivery prices or other incentives associated with the service options that are displayed when customers arrive over time. Such incentives can stimulate efficient fulfillment operations and maximize revenue in the short term.

We present the literature for operational pricing in three clusters, based on the attributes displayed in the lower part of Table 2.2. Even across clusters, the available operational pricing models have many aspects in common. Intuitively, each of them accounts for joint assortment decisions and some form of customer choice behavior. We especially highlight the work of Yang et al. (2016) who calibrate an MNL choice model based on a large amount of real booking data from an e-grocer. Many subsequent publications refer to this model and its data to capture customer choice behavior. Other common features among operational pricing approaches are the use of revenue-based metrics (revenue or profit) for order value assessment and accounting for order delivery as well as opportunity costs in the order assessment. These characteristics largely correspond to those of the second operational offering cluster, which also focuses on the anticipatory steering of customer choice. Within this overall picture, we identify three clusters of publications that differ in terms of the *fulfillment process* design and the method for the *fulfillment feasibility* assessment.

Periodic Fulfillment with Tentative Route Plans Similar to offering, the vast majority of the operational pricing literature assumes a periodic fulfillment process. Within this relatively homogeneous group, the approaches differ mainly in the way they determine fulfillment feasibility. The papers in the first cluster perform a tentative route planning, using insertion heuristics. The tentative route information is also used to estimate profits for assessing the present order value – with or without considering opportunity costs.

Campbell and Savelsbergh (2006) do not consider opportunity costs but estimate the profit contribution of a given order as the sales margin minus the insertion cost, taking into account already accepted customers. An incentive optimization model then trades off price discounts against the increased likelihood that customers will choose time windows with higher profit expectations. More recent approaches seek to also capture opportunity costs, i.e., the impact of demand management decisions on future demand (management). To this end, they typically model the decision problem as a stochastic dynamic program. Yang et al. (2016) are the first to present such a formulation, taking into account the fulfillment costs incurred in the order delivery step. Since this problem is computationally intractable, the authors propose an approximation to compute optimal prices for feasible options in real time. Similar to Campbell and Savelsbergh (2006), the approximation relies on insertion cost estimates, which are offset against the immediate profit before fulfillment. However, the authors incorporate estimates of future demand as they draw on pools of route plans that involve already existing orders and samples of expected future order locations. Koch and Klein (2020) replace the anticipatory insertion cost by a linear VFA that uses the information retrieved from tentative route planning as features. While the former method can only account for cost-related effects in the opportunity cost estimation, this one accounts for both cost- and revenue-related displacement effects. Instead of applying statistical learning, R. Klein et al. (2018) choose a model-based approach to capture these effects. Their MIP formulation combines myopic insertion costs derived from tentative route plans with anticipatory seed-based routing that draws its information from a choice-based demand prediction model.

A major challenge in using tentative route information is computational complexity: The insertion cost calculation is a primary bottleneck (Yang et al., 2016), and it may be necessary to periodically recalculate opportunity costs to decrease online computation times (R. Klein et al., 2018).

Periodic Fulfillment with Capacity Approximation The second operational pricing cluster relies on static capacity controls to assess feasibility instead of using tentative route plans. Alternatively, they skip the feasibility checks altogether and incur penalty costs on capacity shortage. The papers use different approaches to capture the routing aspects of the order delivery step. In addition, they differ in how they link the approximation method used for feasibility assessment to the method used to assess the present order value – in terms of profit or revenue.

Asdemir et al. (2009) and Lebedev et al. (2021) study the structure of an optimal pricing policy under MNL customer choice, assuming static capacity controls. Asdemir et al. (2009) assess the present order value using a revenue metric assuming sunk fulfillment costs. They introduce a balanced capacity utilization constraint to implicitly model the order delivery step. Lebedev et al. (2021) account for delivery costs in the terminal state of their dynamic programming formulation and refer to route approximation methods (C. F. Daganzo, 1987) to determine the assumed capacity controls. The studies show that optimal delivery prices increase dynamically as fulfillment capacities are depleted during order capture (Asdemir et al., 2009), and are monotonic in the number of accepted customers (Lebedev et al., 2021).

The other work in this cluster presents solution methods to the operational pricing problem that involve capacity approximation. Yang and Strauss (2017) build their solution method around C. F. Daganzo (1987). Specifically, they use this approximation method not only to determine static capacity controls for feasibility assessment, but also to train an affine VFA to anticipate profit based on the current number of accepted customers and the time remaining for order capture. A. Strauss et al. (2021) incorporate a similar feasibility assessment but tailor it to a setting with flexible time windows. In particular, they consider a setting in which customers select multiple delivery time windows that are acceptable to them. The customer receives a discount for providing the service provider with more flexibility in order fulfillment. The authors estimate profit through an anticipatory linear program that uses the capacity information from the approximate feasibility assessment. In contrast, Vinsensius et al. (2020) completely ignore feasibility checks at the order capture phase. Instead, they account for infeasibilities in order delivery by means of penalty costs. Yet, the authors incorporate routing properties faced during order delivery: Similar to Yang and Strauss (2017), they estimate profits using VFA. However, rather than relying on approximations, they train their VFA with solutions to a VRP variant with service choice. In particular, they perform the training on simulated historical data and solve the VRP instances using a minimum regret construction heuristic. Thus, although the authors apply explicit route planning within the offline training, they do not perform tentative route planning during the decision-making process, as for example Koch and Klein (2020) do.

Order-based Fulfillment Analogous to operational offering, operational pricing literature addressing order-based fulfillment is scant. In contrast to periodic order fulfillment, delivery decisions are dynamic and stochastic. In what follows, we point out how papers in this cluster deal with this aspect. We also explain how they use tentative route information for assessing opportunity costs. Interestingly, different from the cluster of order-based operational offering literature, none of the considered papers takes order assembly into account.

M. W. Ulmer (2020) dynamically set prices for one-hour and four-hour delivery deadlines. Their model optimizes both the pricing strategy and dynamic route dispatch times, where the former aims to maintain fleet flexibility while charging customers according to their expected willingness to pay. The solution method uses tentative route information obtained from an insertion heuristic that is based on already existing orders only. Besides facilitating feasibility checks, the myopic route information is used to derive fleet flexibility measures as features for a linear VFA that assists profit anticipation. Prokhorchuk et al. (2019) extend this work and aim to make pricing decisions for reliable service assortments to reduce the number of missed deadlines and increase long-term customer loyalty. To this end, they integrate penalties for late deliveries and account for stochastic travel times that materialize while delivery routes are executed. Similar to the above study, the authors build on myopic route information and apply a linear VFA using flexibility- and reliability-based features for anticipatory profit estimation. In contrast, V. Klein and Steinhardt (2023) apply a more advanced tentative routing procedure and consider future orders in both profit estimation and route planning. Compared to previously applied insertion heuristics in combination with route-based VFA, the authors perform a sample-scenario state value approximation that involves heuristically solving a profitable multi-trip VRP with release and due times for every sampled scenario.

2.4. Tactical Demand Management

Table 2.3 lists the literature on tactical offering (upper part) and tactical pricing (lower part). Similar to the previous section, we categorize the publications based on their *problem setting*, the *decision-making process*, and the *computational study*. However, the attributes considered within each of these categories differ from those used to structure the operational literature. Again, the table entries allow us to identify clusters of closely related publications, which we discuss in Sections 2.4.1 and 2.4.2.

First, we distinguish different problem settings underlying tactical demand management in terms of the number of *service options* from which an individual customer can choose (single or multiple) and the *service segments* for which different offering and pricing decisions are made (individual customers or aggregated customer groups).

Second, we consider the forecast-based, tactical decision-making process. Corresponding demand management methods apply different optimization approaches and demand forecasting methods. Optimization approaches differ in terms of the linkage between *planned shifts*, i.e., they determine the decisions either independently for single shifts or jointly for multiple shifts. Further, we distinguish different *model decisions*, including assortment decisions, price decisions, and availability controls. While assortment decisions assign sets of service options to the given service segments, availability controls (e.g., booking limits) are set for given assortments with the aim of simplifying subsequent operational decisions. Finally, we list the *model objective* (cost, revenue, or profit) and the type of service and capacity constraints, if any. In the case of a cost objective, *service constraints* ensure an exogenously imposed service level with respect to the number of service options (frequency), the distribution of service times (balance), or subsets of service options that can be either continuous (interval) or discrete (candidates). *Capacity constraints* capture the necessary fulfillment operations and are represented by continuous approximation models (CA), simulation (SIM), or routing models that can be either explicit (ROUTE) or seed-based (SEED). Note that for each paper only one of the possibly several methods applied is given in the table. Concerning the demand forecast, we distinguish between a deterministic and stochastic *demand model* and indicate whether papers explicitly model customer *choice behavior* based on random utility theory (RUT). Other papers do not model customer choice but assume demand to be independent of the service offering.

Third, analogous to the operational planning models, information on the computational study includes the type of *demand data* (synthetic or empirical) and the *business sector* of the motivating application.

2.4.1. Tactical Offering

Tactical offering decisions determine the availability of service options before the order capture step. In other words, they allocate the corresponding fulfillment capacity to different service segments, based on demand forecasts. In the upper part of Table 2.3, we observe three clusters of publications that share similarities with respect to the considered *service segments* and *model decisions*. As discussed below, each of the clusters represents a specific planning task within the domain of tactical offering – from the simplification of short-term operational planning to service differentiation and long-term customer agreements.

Table 2.3.: Tactical demand management

		Problem setting		Decision-making process							Computational study	
				Optimization approach					Demand forecast			
		Service options	Service segments	Planned shifts	Model decisions	Model objective	Service constraints	Capacity constraints	Demand model	Choice behavior	Demand data	Business sector
Offering	Cleophas and Ehmke (2014)	Multiple	Aggregated	Indep.	Avail. controls	Revenue	–	SIM	Deterministic	–	Synthetic	E-grocery
	Visser and Savelsbergh (2019)	Single	Aggregated	Indep.	Avail. controls	Revenue	–	ROUTE	Stochastic	–	Synthetic	E-grocery
	N. Agatz et al. (2011)	Multiple	Aggregated	Indep.	Assortment	Cost	Frequency	CA	Deterministic	–	Empirical	E-grocery
	Hernandez et al. (2017)	Multiple	Aggregated	Joint	Assortment	Cost	Frequency	ROUTE	Deterministic	–	Synthetic	Retail
	Bruck et al. (2018)	Multiple	Aggregated	Joint	Assortment	Cost	Balance	ROUTE	Stochastic	–	Empirical	Service
	Côté et al. (2019)	Multiple	Aggregated	Joint	Assortment	Cost	Frequency	ROUTE	Stochastic	–	Empirical	Retail
	Mackert (2019a)	Multiple	Aggregated	Indep.	Assortment	Profit	–	SEED	Deterministic	RUT	Synthetic	E-grocery
	Spliet and Gabor (2015)	Single	Individual	Indep.	Assortment	Cost	Interval	ROUTE	Stochastic	–	Synthetic	B2B
	Spliet and Desaulniers (2015)	Single	Individual	Indep.	Assortment	Cost	Candidates	ROUTE	Stochastic	–	Synthetic	B2B
	Spliet et al. (2018)	Single	Individual	Indep.	Assortment	Cost	Interval	ROUTE	Stochastic	–	Synthetic	B2B
P	R. Klein et al. (2019)	Multiple	Aggregated	Indep.	Price	Profit	–	SEED	Deterministic	RUT	Synthetic	E-grocery

Availability Controls The first cluster focuses on establishing availability controls for a given assortment of service options, i.e., thresholds that guide the decision on the availability of service options for different service segments. This simplifies operational decision-making and resembles the concept of allocation planning in supply-constrained production planning (Meyr, 2009).

In this vein, Cleophas and Ehmke (2014) propose an iterative algorithm to allocate the fulfillment capacities of a geographically differentiated service assortment to value-based customer groups. They first simulate the order capture phase based on historical booking data and by applying customer acceptance rules from the literature (Ehmke & Campbell, 2014). From the simulation results, they derive booking thresholds for each time window and delivery area. The authors then refine the thresholds for discrete order value buckets using the expected marginal seat revenue (EMSR) heuristic, a classical revenue management tool (Belobaba, 1987). The computational results show that the proposed method can generate significant revenue gains in the case of heterogeneous order values. In contrast, Visser and Savelsbergh (2019) focus on foresighted delivery routes to maximize the generated revenue. Inspired by Dutch e-grocer Picnic, which offers a single time window per day for each delivery area, they present an approach to (i) determine the specific time window to offer in each area and (ii) establish an operational control mechanism to determine when time windows should be closed. Both decisions are guided by a priori routes that are constructed over a set of delivery points with known order volumes and revenues. Order placement and order sequence are uncertain. The authors develop a two-stage stochastic program, where routes are determined in the first stage and generated revenue is simulated in the second stage. To reduce complexity, the study assumes a single vehicle, thereby turning the routing problem into a traveling salesperson problem (TSP). The study presents insight into the structure of optimal a priori routes.

Assortment Decisions for Aggregated Customer Groups Papers in the second cluster determine an assortment of service options for each geographical area within the service region. In particular, by differentiating the assortment over different areas, the service provider can spatially cluster demand but also temporally sequence the clusters to facilitate efficient delivery routes.

In this light, N. Agatz et al. (2011) determine the service assortment per shift across days for different geographic areas. They assign a fixed number of service options out of a given pool of options to each service area with the objective of minimizing the expected fulfillment cost. To decompose the problem per shift, the authors assume weekly demand to be evenly distributed over the service assortment. Additionally, expected demand is known and independent of the service assortment. The paper proposes two solution approaches, one based on continuous approximation (C. F. Daganzo, 1987) and the other based on integer programming. The authors evaluate the resulting assortments by simulation on the operational level and based on real demand data. The results show a reduction in delivery costs compared to uniform assortments, which is most significant if delivery capacity allows a vehicle tour to span several time windows. Mackert (2019a) extend the integer programming-based method with a finite-mixture customer choice model that accounts for heterogeneous revenues and preferences. Furthermore, they eliminate the specification of exogenous service requirements by moving from cost minimization to profit maximization. The authors linearize the choice-based MIP to apply a standard solver and propose a decomposition heuristic for large instances. The computational results confirm that incorporating customer choice behavior can increase profits. The effect is amplified when preferences are more heterogeneous. The authors also investigate the impact of predefined service requirements on profit and find that an inadequate specification can reduce profits. Hernandez et al. (2017) consider independent demand but account for interdependencies between service assortments over consecutive days. Thus, the assortment decision does not decompose by shift, and the authors use a periodic vehicle routing approach to assign weekly assortments to geographic areas. Routes are modeled at the aggregated area level rather than at individual customer locations. The computational study focuses on the performances of two tabu search-based solution methods, which are also compared to an exact solution method.

In another subset of papers, uncertainties in demand forecasts are explicitly considered. Bruck et al. (2018) discuss the business case of an Italian gas provider that cannot apply operational demand management but must ensure service to all customers at regulated prices. The authors make assortment decisions by assigning capacities (i.e., technicians) to a given pool of time windows and ensure service quality by balancing the assortment over all the days of an operating week. The customers' time window choice is uncertain yet independent of the assortment offered. The authors incorporate the stochastic choice in a simulation stage that is part of a two-stage stochastic program. Combined with a multi-depot multiple TSP, this stage enables the evaluation of first-stage assortment decisions. Using real-life booking instances of the industry partner, the authors demonstrate that their method reduces delivery and penalty costs compared to the company's manual process. Côté et al. (2019) extend the degree of uncertainty to customer locations, basket sizes, and service times. They evaluate an assortment's delivery and penalty costs in the second stage of a two-stage stochastic program using a vehicle routing approach that accounts for multiple interrelated periods. The authors perform a computational study on real instances of a Canadian retail company, the results of which show the effectiveness of their method, which outperforms the manual solution obtained by the company.

Assortment Decisions for Individual Customers The third cluster is concerned with the assignment of single service options to individual customers, which can be interpreted as long-term customer agreements – a special case of service differentiation. The set of customers is fixed and known in advance, and all customers have to be served.

Spliet and Desaulniers (2015) and Spliet and Gabor (2015) consider a business-to-business (B2B) case inspired by a Dutch retailer. In this context, 'customers' refer to retail stores that are replenished periodically. The supplier assigns to each store a time window in which it will receive deliveries. This assignment decision is driven by stochastic demand volumes. The authors present a two-stage stochastic linear program that evaluates assignment decisions based on a vehicle routing model. The objective is to minimize delivery costs subject to the stores' preferred delivery time intervals (Spliet & Gabor, 2015) or candidate options (Spliet & Desaulniers, 2015). Both formulations are solved to optimality using a branch-and-price-and-cut algorithm with route relaxations. In subsequent work, Spliet et al. (2018) add time-dependent travel times and seek arrival time consistency. The authors propose an exact solution method and evaluate its performance.

2.4.2. Tactical Pricing

We define tactical pricing as the planned differentiation of prices across both customer groups (e.g., by geographic location or order value) and service options (e.g., premiums for evening delivery). While tactical offering limits an assortment's breadth, tactical pricing steers customers to favorable options within a (potentially broader) assortment. As seen in the lower part of Table 2.3, we are aware of one single publication focused on tactical pricing.

R. Klein et al. (2019) consider price differentiation between time windows offered in given geographic areas, with the objective of maximizing total profit. Assortments are fixed, but prices can be selected from a finite price list. Akin to the majority of operational pricing studies, the authors explicitly model customer choice behavior based on random utility theory. Specifically, they apply a non-parametric rank-based model that captures a customer segment's choice behavior through preference lists over all possible service options, including non-purchase. The authors formulate the pricing problem as an MIP that either features aggregate vehicle routes or cost approximations with respect to the geographic areas. The computational results confirm the benefits of differentiated pricing over uniform pricing. For industry-sized instances, the authors recommend their approximation-based approach since it is able to find good solutions in a limited amount of time.

2.5. Strategic Demand Management

The studies on operational and tactical demand management discussed in the preceding sections make assumptions regarding the setting defined by strategic-level decisions. These include decisions on the service region, appropriate service segments, the service design, and the pricing model. Interestingly, publications that address these decisions in their own right are few and far between. Therefore, rather than creating a literature table similar to those in Sections 2.3 and 2.4, we present the problem- and methodology-related focus of the current state-of-the-art literature on strategic demand management at a glance in Table 2.4. We discuss the relevant aspects of key strategic planning tasks and contextualize current perspectives in the literature. As in the preceding sections, we distinguish between offering and pricing levers.

2.5.1. Strategic Offering

Strategic offering refers to identifying target markets and designing an appropriate service proposition, which translates to three major planning tasks that guide our discussion: The selection of the service region, service design, and the definition of service segments (see Roth & Menor, 2003).

We start with the literature that sheds light on the choice of *service region*. Here, a decision has to be made whether to offer service in a densely or sparsely populated area.

Table 2.4.: Strategic demand management

		Planning task				Main methodology
		Service region	Service design	Service segments	Pricing model	
Offering	Lin and Mahmassani (2002)	✓	✓			Simulation
	Wilson-Jeanselme and Reynolds (2006)		✓			Empirical
	Boyer et al. (2009)	✓	✓			Simulation
	M. Ulmer (2017)		✓			Simulation
	Manerba et al. (2018)		✓			Scenario evaluation
	Ramaekers et al. (2018)	✓	✓			Scenario evaluation
	Amorim et al. (2024)		✓			Empirical
	Bruck et al. (2020)			✓		Prescriptive
	Milioti et al. (2020)		✓			Empirical
	Fikar et al. (2021)		✓			Simulation
	Magalhães (2021)		✓			Empirical
	Phillipson and Van Kempen (2021)		✓			Simulation
	Rodríguez García et al. (2022)		✓			Case study
Pricing	Gümüş et al. (2013)				✓	Game-theoretic
	Belavina et al. (2017)	✓			✓	Game-theoretic
	N. Agatz et al. (2021)				✓	Simulation
	Wagner et al. (2021)				✓	Prescriptive

The former includes mostly metropolitan areas and inner cities with dense road networks and high demand potential but also more fierce competition. The latter is characterized by sparser road networks and lower customer density but may allow the retailer to achieve a monopoly. In this vein, several studies have examined the operational implications of urban and rural service regions (Belavina et al., 2017; Boyer et al., 2009; Lin & Mahmassani, 2002; Ramaekers et al., 2018) and conclude that customer density has a significant positive effect on route efficiency. Beyond strategic demand management literature, Jiang et al. (2019) discuss general challenges of last-mile delivery in rural, more sparsely populated areas. In the operational demand management literature, Ehmke and Campbell (2014) and Köhler and Haferkamp (2019) show that the characteristics of the service region also influence which real-time order evaluation method is most appropriate.

Second, we consider the literature addressing *service design*. This planning problem refers to a broad spectrum of design elements that characterize a delivery service offer and its service level. This includes decisions on delivery speed (e.g., click-to-door time), precision (e.g., time window length), and service frequency. Further design decisions concern possible interactions between service assortment and physical assortment, customer flexibility in terms of changes in the time window and shopping basket, and value-added services such as returns management. To gain a competitive advantage, it is important to understand both the sales impact and operational implications of different service designs (Amorim et al., 2024). Thus, on the one hand, many empirical studies have investigated customer preferences and expectations regarding particular delivery service attributes (Amorim et al., 2024; Magalhães, 2021; Milioti et al., 2020; Wilson-Jeanselme & Reynolds, 2006). Most recently, Rodríguez García et al. (2022) present a framework on how to map value proposition to logistics strategy, thereby qualitatively assessing operational implications of a service design. All of these studies shed light on how service design attributes affect the generated demand volume.

On the other hand, there is a wide field of exploratory research that examines the operational implications of a service design. Starting in the early 2000s, Lin and Mahmassani (2002) show by simulation that increasing the time window length can reduce vehicle idle time, lower total miles traveled, and allow for more customers to be served. Boyer et al. (2009) support their results, and Ramaekers et al. (2018) report similar effects for both delivery and assembly operations. M. Ulmer (2017) focus on the impact of offering delivery deadlines, and Manerba et al. (2018) investigate both click-to-door time and time window length from an environmental perspective. N. Agatz et al. (2011) perform a sensitivity analysis on the choice of service frequencies, and Mackert (2019a) show that an inadequate specification can reduce profits. Very recently, Phillipson and Van Kempen (2021) have assessed the cost implications of allowing customers to change their chosen time window before the delivery day, and Fikar et al. (2021) have examined the integration of product shelf-life options into demand management decisions. Some of these findings have already been picked up in operational demand management: Casazza et al. (2016) perform dynamic service design adjustments, and Campbell and Savelsbergh (2006) and Köhler et al. (2020) offer and price time windows depending on their length.

Lastly, we present literature that concerns defining appropriate *service segments* which form the basis for tactical service differentiation. It should be noted that these segments do not necessarily coincide with the customer segments used to capture different preference structures within customer choice models. Tactical demand management commonly assumes given service segments based on geographic characteristics such as a customer's zip code affiliation; only Cleophas and Ehmke (2014) additionally group customers based on their basket value (see Table 2.3). We are aware of just a single contribution that determines optimal service segments in this context. Bruck et al. (2020) extend the tactical approach of Bruck et al. (2018) and integrate strategic offering. They determine optimal service segments by solving a P-median facility location problem to group municipalities within the considered service region. A service constraint handles potential imbalances among segments' total expected demand. The authors evaluate their approach using real industry data and emphasize its value for assessing entry into new service regions and analyzing past service segment configurations.

2.5.2. Strategic Pricing

Strategic pricing refers to the overall pricing model and depends on the competitive environment, customer preferences, and price sensitivities within the target market. Determining a pricing model includes decisions about free or paid delivery, whether to use a delivery charge per order or a subscription fee per service period, and other incentive schemes. Tactical and operational demand management commonly assume a per-order pricing model within a given price range to steer customer choice. However, we are aware of several studies that shed light on the impact of specific pricing models.

Belavina et al. (2017) consider grocery delivery and build a stylized model to examine per-order and subscription-based pricing models with respect to equilibrium customer behavior and resulting profit and environmental performance. Their results show that subscription-based models lead to more frequent delivery requests, which in turn impact the provider's revenue, route efficiency, and food waste. The authors conclude that the subscription model tends to be more environmentally friendly because the reduction in food waste emissions outweighs the increase in delivery emissions, but they still recommend the per-order model for high-margin providers that operate in sparsely populated areas. Wagner et al. (2021) show that on average, the increased order frequency entails a profit loss as the increase in assembly and delivery costs outweighs the increase in revenue. The authors explain this effect as a result of higher expectations of subscription customers; i.e., they choose narrower and more popular time windows. In addition, the authors develop a data-driven algorithm that predicts the expected post-subscription profitability to determine whether a particular customer should be offered a subscription plan. The algorithm is trained and evaluated based on real order data from a large omnichannel grocery retailer. The authors report that observed product assortment size and basket value are the strongest predictors of post-subscription profitability. In contrast, Gümüş et al. (2013) investigate the joint design of a pricing model for product and delivery service. They analyze the competitive dynamics of price partitioning, where delivery and product prices are displayed separately in a partitioned setting, and free shipping is advertised in a non-partitioned setting because the delivery cost is already included in the product price. The authors determine the equilibrium market structure and validate their theoretical results through empirical analyses. In addition to traditional pricing models, N. Agatz et al. (2021) focus on non-monetary incentives and study the impact of displaying green labels for environmentally friendly service options on customer behavior and operational performance. From their empirical experiments and simulation study, the authors verify that green labels effectively steer customer choice, also in combination with price incentives and for less attractive time windows.

2.6. Discussion

In this section, we synthesize our findings from reviewing the literature, highlight key challenges and potential future research for each planning level, and elaborate on the connection between the planning levels.

There is a growing number of academic contributions on *operational* demand management, predominantly directed at e-grocery. The computational challenges make it an active field of research in operations research. Most work in this area focuses on sophisticated solution methods for specific parts of the real-time decision problem, e.g., feasibility assessment, value anticipation, or customer choice behavior. In general, vehicle routing heuristics and dynamic programming can be identified as methodological cornerstones.

Building on the current body of research, we see several avenues for future research. First, given the modular structure of operational decision-making, there is a need for comprehensive benchmarks that guide the selection of suitable building blocks of solution methods. Lang and Cleophas (2020) and M. W. Ulmer (2019) offer valuable starting points for this purpose. Second, in light of very limited computation time, there is still a need for fast solution methods. One potential research avenue is the application of machine and reinforcement learning in this context. Such methods have already been adapted for feasibility assessment (van der Hagen et al., 2022) and value anticipation (e.g., Koch & Klein, 2020) but have not yet been applied to predict customer choice. Alternatively, it may be beneficial to change the fulfillment process design to simplify operational planning. We see valuable starting points in the recent literature: Schwamberger et al. (2023) define an inverted order capture process in which the service provider proactively approaches customers with the opportunity to place an order, and Yildiz and Savelsbergh (2020) explore the possibility of incentivizing accepted customers to change their chosen time window after the order capture cut-off time.

We see fewer contributions to *tactical* demand management that, however, cover a variety of planning problems from long-term customer agreements to short-term availability control. From a methodological perspective, MIP, two-stage stochastic programming, and simulation are prevalent and customer choice behavior is rarely modeled explicitly. Besides, we observe that tactical approaches are mainly tailored to specific business sectors and that the research is often conducted in collaboration with an industry partner, which indicates the practical relevance of the topic.

We see a need for future research, especially for innovative AHD concepts. Service providers that perform order-based fulfillment within a deadline benefit from tactical offering and pricing decisions: Different delivery deadlines can be offered in different geographic areas at different prices (e.g., longer and/or more expensive deadlines in peripheral areas). Stroh et al. (2022)'s tactical vehicle dispatch policies may serve as a starting point. Moreover, there is great potential for tactical offering under a subscription-based pricing model. Spliet and Desaulniers (2015), Spliet and Gabor (2015), and Spliet et al. (2018) provide relevant insights from the business-to-business context that can be transferred to customers who are allowed to reserve a time window as part of their subscription plan.

Contributions to *strategic* demand management provide insight into many different aspects of strategic planning. The set of applied methodologies is much more diverse which we explain by the strong interdependencies with other domains. For example, selecting a service region interacts with location planning, determining service segments is influenced by delivery districting (e.g., Banerjee et al., 2022; Haugland et al., 2007), and service design and pricing models strongly depend on marketing and competitive considerations. As a consequence, we see that comprehensive decision support is still missing. Other reasons that might promote this gap are that (i) strategic demand management decisions are considered to have less leverage compared to strategic decisions in other research fields (e.g., network design) since they are less long-term and more easily reversible. (ii) Competitive constraints may leave only limited room for optimization. (iii) From a practitioner's perspective, decision-making responsibilities are more dispersed and located at a higher managerial level than they are for tactical and operational demand management.

We see the opportunity for strategic demand management to provide comprehensive decision support to capture the greatest possible demand potential and to do so profitably. Thereby, important issues of competitive pressure and market share should also be addressed. Looking to adjacent research fields confirms this potential. Metters and Walton (2007) provide strategic decision support by proposing a service sector typology for multi-channel e-tailing. They develop a matrix of competitive positions along the dimensions of inventory pooling and shipping consolidation, and identify four types of strategies that can be adopted by multi-channel e-tailers. The authors also emphasize that e-tailers should align their supply chain configuration with their strategic objectives. For the express delivery business sector, F. Li et al. (2021) propose a two-dimensional decision matrix to select the most suitable delivery service mode among direct and indirect options. They measure the expected customer utility and calculate the expected cost of delivery service to map different service modes to the decision matrix.

We conclude our discussion with a few observations concerning the *interaction* between the different planning levels reviewed separately in Sections 2.3–2.5. Conceptually, longer-term decisions set the boundaries for decisions on the shorter term. One challenge is that actual performance can only be observed once orders materialize. Appropriately anticipating this performance impact is a core issue for long-term decisions. Given the scarcity of strategic demand management research highlighted in Section 2.5, the impact of corresponding long-term decisions on tactical and operational demand management is largely an open issue to date. Most contributions to the tactical and operational literature make assumptions on the strategic system design, based on choices observed in practice. However, the appropriateness of these choices, including the service region, service design, and service segments has received limited attention thus far.

As a potential starting point for future research in this direction, some studies consider the sensitivity of tactical or operational decisions and their performance to changes in selected strategic choices. Examples are strategic choices between suburban and downtown service regions (e.g., Ehmke & Campbell, 2014) and between different time window lengths (e.g., Campbell & Savelsbergh, 2005; Côté et al., 2019). Conceptually, these studies follow a what-if approach to strategic-level decisions. A next step would be to turn the analysis into a systematic optimization approach that selects strategic options based on their impact on day-to-day operations and performance. For example, N. Agatz et al. (2021) conducted operational-level simulations to assess the potential of new ways for steering customer behavior. Their strategic concept of green labels can be incorporated in tactical and operational pricing, complementing the current monetary incentives.

Interactions between the tactical and operational planning levels have received more attention in the literature. This is primarily driven by the fact that operational demand management decisions must be made in real time to facilitate a smooth order capture process. This limits the available time for computations on the operational level. There is, however, more time to support tactical decisions. We observe two approaches in the literature that exploit this relation.

First, tactical decisions can pre-structure and thereby simplify operational decisions by limiting the decision space on the operational level. In the reviewed literature, this holds true for service and price differentiation. To be effective, such approaches must capture the link with the operational level. The extent to which this is the case depends on the decision-making flexibility assumed at this level. Long-term service agreements (e.g., Spliet & Gabor, 2015), legal regulations (Bruck et al., 2018), or business policies (Côté et al., 2019) may severely limit operational levers. In these cases, we observe more accurate routing formulations and the use of two-stage stochastic programs to hedge against forecast errors. If, on the other hand, operational demand management opportunities are more extensive, the demand model and operational impacts are more coarsely estimated (N. Agatz et al., 2011; Hernandez et al., 2017; R. Klein et al., 2019; Mackert, 2019b). However, operational performance may be tested outside of the decision model, through simulation studies (e.g., N. Agatz et al., 2011).

Second, it may be beneficial, or even necessary, to shift some decisions from the operational to the less time-constrained tactical planning level altogether. Essentially, this implies a choice between an elaborate ex-ante planning model and a simpler heuristic using real-time information. Given the discussed computational limits, it makes sense to reserve real-time planning to those decisions for which the available real-time information really makes a difference. One example of shifting decisions to the tactical level is the ex-ante calculation of availability controls such as booking limits for specific time windows (Cleophas & Ehmke, 2014). Corresponding literature uses simulation and two-stage stochastic programming to capture the effects on the operational level (Cleophas & Ehmke, 2014; Visser & Savelsbergh, 2019).

2.7. Conclusion

This review paper introduced a framework for classifying demand management decisions for AHD with respect to different planning levels and demand management levers. For each planning level, we presented and classified prescriptive analytics methods in the literature and identified research gaps. The following are our main observations. We have seen a rich set of studies on *operational* demand management, aimed at extracting the greatest potential from real-time decisions. Because manifold opportunities for real-time decision-making differentiate AHD from traditional brick-and-mortar retail, the appeal of this line of research is intuitive. The ensuing computational challenges have triggered sophisticated algorithmic contributions. However, all decisions clearly do not benefit equally from real-time information. In this light, we see yet unlocked opportunities for *tactical* demand management to simplify and prestructure operational decisions. Finally, there is a striking lack of research on underlying long-term, design-level decisions. Hence, we see great potential for future contributions to *strategic* demand management for AHD.

Taking a more general perspective, we highlight four topical themes that we believe hold opportunities for innovative and relevant future research on demand management for AHD. These themes give rise to novel analytics issues at all planning levels.

First, a natural direction concerns innovative business models and services in AHD. While research on standard 'next-day' grocery delivery is maturing, researchers have only started to study new delivery trends. On the one hand, on-demand e-grocery startups (e.g., Gorillas and Flink) promise 'instant' grocery delivery within a few minutes. This fundamentally different service offering challenges many assumptions of the current fulfillment strategies and corresponding demand management. On the other hand, established businesses are exploring novel customer interaction processes that deviate from the current standard process reflected in Section 2.2.1. Examples include long-term subscription agreements and proactive customer contacting. These developments give rise to novel decisions and call for corresponding analytics models and approaches.

Second, more research that addresses new objectives in demand management for AHD is needed. To date, the majority of publications focus on profit maximization as the primary goal of service providers. Given the expansion race between emerging on-demand e-grocery businesses, research should recognize market share as a relevant alternative objective. Furthermore, considering environmental objectives has become a standard in many research fields, and delivery services are subject to particular public scrutiny with regard to sustainability (Siragusa & Tumino, 2021). Belavina et al. (2017) and Manerba et al. (2018) are the first to investigate the leverage of demand management in light of environmental objectives. Future research should expand this development and explore the impact of multiple conflicting objectives, for example, related to social responsibility toward internal stakeholders (e.g., delivery workers) and external stakeholders (e.g., customers, residents, and administrators). Recent literature has underlined the relevance of this perspective: Belanche et al. (2021) show that customers' purchase intentions depend on their perception of the working conditions for delivery workers, Chen et al. (2023) and Soeffker et al. (2017) investigate demand management regarding fairness to customers, and Bjørgen et al. (2021) discuss the integration of e-grocery logistics into urban spaces. The rapid expansion of micro depots to support instant grocery deliveries, so-called 'dark stores', have already sparked public and political debate: The Dutch cities of Amsterdam and Rotterdam recently restricted the opening of new facilities because of noise and the blocking of pedestrian walkways[11].

Third, we see potential for demand management addressing the interaction between the delivery service and the product assortment. Fikar et al. (2021) and Gümüş et al. (2013) provide initial work in this direction. Future research may strengthen the integration of product assortment-related aspects into demand management and extend demand management levers accordingly. For example, while existing levers have been shown to effectively reserve fulfillment capacity for more valuable customers, the inventory rationing literature demonstrates a similar effect with respect to product availability by reserving inventory for high-margin customers (e.g., Jimenez G et al., 2020). In addition, integrating the product assortment naturally draws attention to the order assembly process. We have seen few contributions that explicitly account for order assembly in demand management methods. Among those is research exploring the impact of time windows on both assembly and delivery (Ramaekers et al., 2018) and research presenting operational offering for order-based fulfillment (Azi et al., 2012; Klapp et al., 2020). Product-related demand management requires new analytical models and approaches that enable integrated decision-making at all planning levels.

Fourth, we call for more empirical validation of demand management for AHD. On the one hand, we recognize that results based on empirical instances alone are difficult to generalize and should therefore be supported by carefully generated synthetic data.

[11]Reuters, https://reut.rs/3HRBLh9. Accessed on February 14, 2022

The classification of demand data presented in Tables 2.2 and 2.3 is intended to shed light on this crucial aspect, even though the observed situation is more nuanced than a strict dichotomy. While research on supply-oriented levers can more easily base the computational results on synthetic instances, empirical data are particularly important for demand management because of the strong role of customer interaction in this context. Many of the assumptions required for demand management relate to customer behavior, which is difficult to model realistically without empirical data. In addition, customer behavior changes over time, so empirical validation should be reviewed regularly.

To conclude, we expect demand management for AHD to continue to gain importance and to witness significant innovations to emerge. We hope that this review contributes to stimulating future research into this dynamic field.

Chapter III

Evaluating Time Window Assortments for Grocery Home Delivery[1]

with Niels Agatz and Moritz Fleischmann

Abstract

Online grocery services require the customer to be present at the time of delivery. The resulting time window constraints pose a major challenge to the profitability of the online grocery business model. In practice, we see great variation in the time window assortment, i.e., the set of time windows offered to customers, including long or short, many or few, overlapping or non-overlapping options. For the success of an online grocery business, it is essential to understand how these choices impact demand as well as delivery efficiency. We develop a model to evaluate time window assortments in terms of these performance metrics. The evaluation model can incorporate different types of demand functions, and it approximates the components of the delivery system through tractable functional expressions using continuous approximation. In our study, we identify and analyze fundamental trade-offs underlying the time window assortment. Our analytical results provide insights that are instructive for designing time window assortments, thus helping practitioners to align their operations strategy with relevant market conditions to become economically viable.

[1]The contents of this chapter build on the working paper Waßmuth et al. (2025).

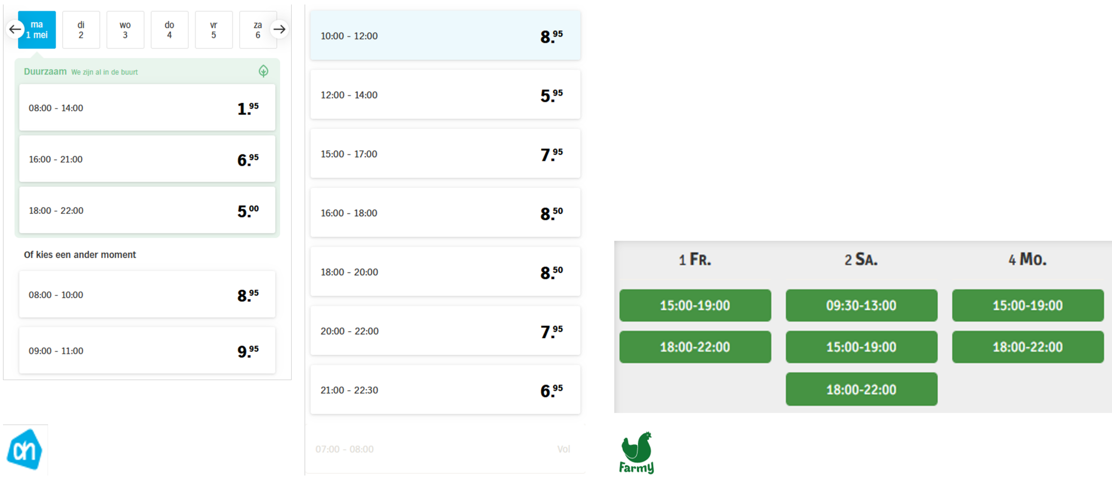
3.1. Introduction

Online grocery sales continue to grow globally, albeit at a slower pace than during the COVID-19 pandemic (McKinsey, 2022). Because groceries are perishable and bulky, the customer needs to be present at the time of delivery, a concept known as *attended home delivery*. To minimize missed deliveries and customer waiting time, service providers commonly use time windows to align delivery times with their customers. More specifically, the service provider offers a menu of time windows from which the customer selects one. For the service provider, this raises the question of which time windows to offer in the first place – an issue which requires decisions on the number of time windows, their lengths, timing, and potential overlap. These decisions shape the company's service offering, analogous to assortment decisions for physical products. Emphasizing this analogy, we refer to the above decisions as time window *assortment* decisions.

In developed economies, companies struggle to make online grocery delivery profitable, due to logistically demanding products, low margins, and high labor costs. These challenges require a careful cost-benefit trade-off between demand effects and supply efficiency. Time window management impacts both sides of this trade-off and is therefore a crucial profit lever. While time window management is an active research field, most of the available literature takes a *given* set of time windows as input and focuses on whether or not to offer a particular window to a particular customer (Fleckenstein et al., 2023; Waßmuth et al., 2023). How to compose the initial set of time windows, i.e., the time window assortment, is not yet well understood.

The significance of this decision is evident in practice, where companies in various regions with unique geographic and demographic features offer time window assortments that vary noticeably in both the number and length of time windows (see Figure 3.1). We aim to contribute to filling this gap by developing a model that allows us to capture and analytically investigate the fundamental trade-offs underlying the time window assortment design.

Planning and execution of grocery delivery operations involves multiple decisions on different hierarchical levels, with time window assortment design as the first step, followed by capacity planning, order capture, and fulfillment planning, as illustrated in Figure 3.2. We focus on periodic fulfillment processes, which means that order fulfillment is organized in *shifts*, each having a certain cutoff time until which orders are accepted. This setup is common for many full-assortment online grocers. In contrast,



(a) Dutch e-grocer Albert Heijn (b) Swiss e-grocer Farmy

Figure 3.1.: Exemplary time window assortments in practice

quick-commerce providers tend to use order-based fulfillment, relying on an individual delivery deadline for each order (Waßmuth et al., 2023).

In periodic fulfillment, the set of customer orders for a given shift is known, once the cutoff time has passed. Thereupon, the delivery routes are planned, orders are picked in the warehouse, and then delivered. These are the processes that drive the fulfillment costs. In this paper, we focus on fulfillment cost effects that occur during the order delivery phase. Revenues, on the other hand, materialize before the cutoff time, when customers place their orders. During this order capture phase, the service provider decides which time windows to offer to a newly arriving customer.

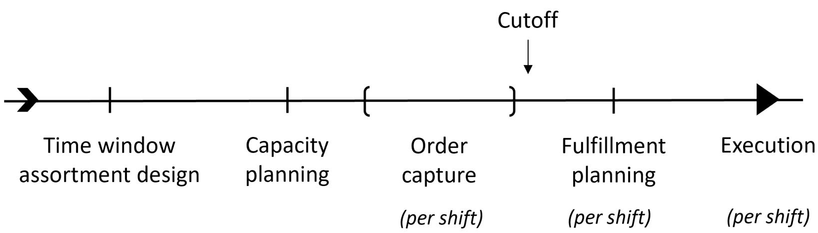


Figure 3.2.: Illustration of a typical planning process in online grocery delivery

Theoretically, the company could try and optimize all attributes of the time window offering specifically for each arriving customer. However, two issues prohibit such a highly individualized service offering. First, the offering decisions must be made almost instantaneously to ensure a smooth order process for the customer. This rules out a large scale optimization over many possible sets of time windows (van der Hagen et al., 2022). Secondly, an overly tailored time window offering may confuse customers, potentially leading to a decline in repeat orders. The time window assortment planning simplifies decisions and communication during order capture by limiting the available options to a structured menu of time windows. In addition, the assortment plan can guide tactical capacity planning, regarding the required fleet and workforce. In conclusion, designing an appropriate time window assortment is a strategic planning task at the sales and operations interface (see Rooderkerk et al., 2023).

We characterize a time window assortment by the length, number, and possible overlap of the time windows (see Cordeau et al., 2023). Time windows can be sequential (9:00-10:00 and 10:00-11:00) or partially overlapping (9:00-10:00 and 9:30-10:30). One can also mix different time window lengths and overlaps, for example, 09:00-10:00, 10:00-11:00 and 9:00-11:00. Empirical research shows that the length of the delivery window (also referred to as 'precision') and the number and timing of the time windows affect customer choices (Amorim et al., 2024). On the other hand, it has long been known that time window length affects operational flexibility in order delivery, and therefore efficiency and costs (Boyer et al., 2009; Lin & Mahmassani, 2002). Moreover, overlapping time windows allow for more time window options or longer time windows within a fixed delivery shift (N. Agatz et al., 2011; Campbell & Savelsbergh, 2005). However, the operational implications remain to be fully understood. Only in terms of problem complexity has it been observed that allowing overlapping time windows increases the size of the corresponding optimization models (N. Agatz et al., 2011; Hungerländer & Truden, 2018; Truden et al., 2022).

In this paper, we develop an analytical model to evaluate the impact of time window assortment decisions on the performance of the delivery system. This model aims to provide guidance in assessing different time window assortments in terms of profitability (revenues and costs) and fleet requirements. Our contribution is threefold: (i) we develop an evaluation model for time window assortments that captures both demand and supply effects, (ii) we identify fundamental trade-offs in time window assortment design and provide analytical expressions to capture them, and (iii) we derive managerial insights to support time window assortment decisions and growth strategies. These insights help practitioners align their operations strategy with market conditions, enhancing their economic viability.

The remainder of this paper is organized as follows. In Section 3.2, we review the related literature. We present the time window assortment evaluation model in Section 3.3, which includes the statement of key assumptions and notation. In Section 3.4, we unveil what factors drive the profitability of a time window assortment as well as the required vehicle capacity. We then use the obtained insights to analyze how assortment decisions about the length of a given number of time windows (Section 3.5) and about the number of time windows of a given length (Section 3.6) affect both demand and delivery performance. We summarize our results and conclude with managerial implications in Section 3.7. All proofs are provided in Appendix A.

3.2. Related Literature

Online grocery retailing is receiving increasing attention from operations management scholars. Recent publications inspired by or applied to online grocery cover a wide range of topics, from omnichannel retail strategy (Delasay et al., 2022) and order fulfillment (Dayarian & Pazour, 2022) to inventory allocation (Feng, Li, et al., 2022) and delivery pricing (N. Agatz et al., 2021; G. Li et al., 2023). Our contribution to this literature builds on three streams of research: time window management, (service) assortment planning, and continuous approximation models for last-mile delivery planning.

Managing the delivery service offering to the customer, commonly referred to as time window or time slot management, is part of a broader area of research on demand management for attended home delivery and has gained increasing interest (Cordeau et al., 2023; Fleckenstein et al., 2023; Waßmuth et al., 2023). Most of this work is motivated by examples in online grocery delivery, but it also applies to other attended home delivery settings, for example as related to service engineers, furniture, and home appliances. The existing literature on time window management is primarily concerned with operational and tactical planning tasks, typically assuming a fixed set of potential time windows as input. This body of work focuses on assigning time windows to customers or geographic regions, either on the basis of actual demand (e.g., Campbell & Savelsbergh, 2005), or on the basis of forecasts (e.g., N. Agatz et al., 2011). In contrast, our work addresses the design of time window assortments as a strategic planning task.

A few contributions report simulation studies that assess the impact of various delivery service configurations, some of which can be transferred to time window assortment decisions, on a range of operational performance metrics. Lin and Mahmassani (2002) present a simulated scenario analysis to examine trade-offs across five key variables, while Punakivi and Saranen (2001) and Boyer et al. (2009) provide simulated response functions, albeit for fewer variables. None of these studies treat demand as endogenous with respect to the delivery service offering. Furthermore, the computationally intensive routing procedures in the simulations restrict the analyses to small sets of variables. In contrast, our paper accounts for endogenous demand and provides analytical expressions that capture the fundamental trade-offs across multiple variables of interest.

Time window assortment design in e-grocery resembles product assortment planning in retail, where demand is endogenous. Assortment planning aims to identify the set of products (or product categories) that maximizes expected revenue (or profit) while adhering to operational constraints such as shelf space limitations. Kök et al. (2015) and Heger and Klein (2024) provide excellent reviews of the existing literature. Work in this area is characterized by a trade-off between accurately modeling customer choice behavior and effectively solving the corresponding assortment optimization problem (Chung et al., 2019). Accordingly, some works focus on modeling customer choice behavior (e.g., Chung et al., 2019; Le & Mai, 2024; Nip et al., 2021; Y. Wang & Shen, 2021) while others focus on solving the assortment optimization problem under a variety of operational constraints (e.g., Feldman & Paul, 2019; Feldman & Topaloglu, 2015; Hübner et al., 2020). While the difficulties in solving assortment problems for physical products are primarily related to the level of sophistication of the demand models, our *service* assortment is intricately linked to the complex underlying routing processes through the delivery time windows. This means that we have to make careful modeling trade-offs about the level of detail with which we incorporate demand models and vehicle routing.

Service assortments have been previously addressed in academic literature, with airline operations serving as a prominent example within the domain of transportation service systems. Airline planning bears resemblance to time window management, primarily due to two key factors. Firstly, the product in question is a service, namely commercial passenger flights. Secondly, decisions concerning, for example, strategic route plans (e.g., Birolini et al., 2021) or tactical flight schedules (e.g., Cadarso et al., 2017; Wei et al., 2020) are significantly influenced by the interplay between passenger demand and aircraft supply. Birolini et al. (2021) tackle a planning problem that shares some similarities with the one addressed in this research. Their work introduces an innovative approach to strategic airline planning, incorporating route selection, flight frequency, and fleet composition. This approach takes into account the interaction between demand and supply, where the set of available flight options impacts passenger demand, which subsequently influences the scheduling of aircraft operations. Compared to our work, the authors employ a distinct methodology, estimating an empirical passenger demand model and subsequently integrating it into an optimization model, which they solve numerically in a case study.

In contrast, we build a continuous approximation model to capture the interactions of demand and downstream vehicle routing. The core concept of the continuous approximation paradigm is to approximate complicated combinatorial quantities using simpler mathematical expressions, based on densities (see C. Daganzo, 2005). Continuous approximation models have been widely applied to various last-mile delivery problems. We refer to Ansari et al. (2018) for a comprehensive review. Our paper aims to evaluate the performance and model the key trade-offs associated with different time window assortment decisions. It does not seek to provide detailed operational plans, but rather focuses on deriving insights from closed-form solutions of stylized mathematical expressions. In this sense, it resembles the works of Belavina et al. (2017), Stroh et al. (2022), and Smilowitz and Daganzo (2007), all of which provide valuable insights for the design and performance evaluation of last-mile delivery systems using simple density functions.

3.3. Assortment Evaluation Model

We consider an e-grocery retailer that lets customers select a delivery time window from a menu of options, which we refer to as the time window assortment $T = (l, n, o)$. This assortment is characterized by the time window length l, the number of time windows n, and a categorical indicator o specifying whether the time windows are consecutive or overlapping.

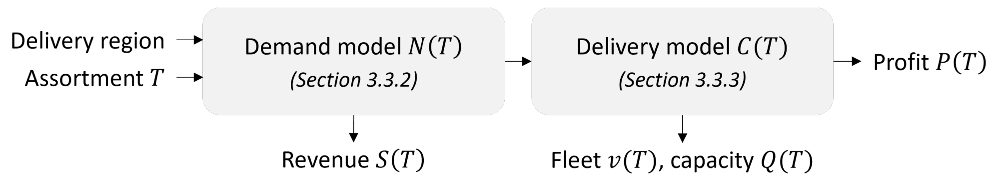


Figure 3.3.: Conceptual representation of the time window assortment evaluation model

In this section, we develop a modeling framework to evaluate the impact of the time window assortment on both customer demand and delivery operations. Figure 3.3 depicts our framework, which consists of a demand model and a delivery model. The demand model specifies choice behavior, customer demand, and revenue in the order capture phase. Given that demand, the delivery model approximates the relevant operational and cost metrics of the order delivery phase. We outline our key assumptions in Section 3.3.1 and then present our demand model in Section 3.3.2 and our delivery model in Section 3.3.3. Table 3.1 summarizes the main notation.

3.3.1. Assumptions

Each time window assortment is associated with a particular delivery shift. Moreover, the retailer operates from a single delivery hub (see Boyer et al., 2009; Punakivi & Saranen, 2001) for a certain circular delivery region. Deliveries are affected by the time window constraints imposed by the time window assortment and the demand for that assortment. Our model aims to capture the major strategic trade-offs in time window assortment design and builds on the following key assumptions.

Assumption 3.1. *Delivery shifts are independent.*

We treat delivery shifts as independent, which implies that there are no interactions between shifts, neither in terms of demand nor delivery operations. This is a common demand-side assumption in the literature on time window management for attended home delivery (Waßmuth et al., 2023) that is empirically supported (Amorim et al., 2024; Yang et al., 2016). The independence of delivery operations between shifts is inherent to periodic fulfillment, since all operations take place within a given shift.

Assumption 3.2. *Time window assortments, customer locations, demand, orders, and vehicle tours are homogeneous.*

This assumption encompasses several parts of the fulfillment process and ensures smooth, uniform conditions, as is typical in the application of the continuous approximation method (C. F. Daganzo, 1987). First, we consider a homogeneous time window assortment, meaning that (i) each time window has the same length, and overlapping time windows overlap symmetrically with adjacent windows – for example, 8:00-10:00, 9:00-11:00, and 10:00-12:00; and (ii) the set of time windows in the assortment is connected, i.e., there are no gaps between consecutive windows. Second, we assume that customer locations are homogeneous, meaning they are uniformly distributed across the delivery region, and that all customers are offered the same time window assortment. Third, homogeneous demand implies that demand is evenly distributed across all time windows in the assortment. Fourth, we assume homogeneous orders, meaning that all customer orders have identical gross sales margins (including product margins from the shopping basket and delivery fees) and require the same vehicle capacity in terms of weight and volume. Lastly, homogeneous vehicle tours imply that each tour extends to the entire delivery shift and serves an identical number of customers.

3.3.2. Demand Model

In our framework, the demand model $N(T)$ offers a projection of the number of orders per time window for time window assortment T. From the demand model, we obtain the total gross margin as $S(T) = rnN(T)$, with r denoting the unit gross margin per order.

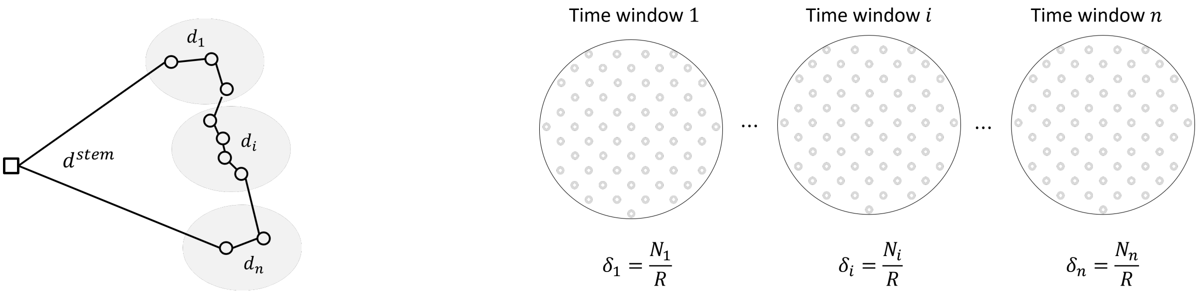
Our modeling framework is flexible and can incorporate any demand model to capture the customer choice behavior for a given time window assortment. In our analytical study, we examine two stylized demand models to gain insights into the key trade-offs associated with substitution behavior. In Section 3.4, we study the impact of demand as an exogenous parameter, N. In Sections 3.5 and 3.6, we model demand endogenously, either as a function of time window length, $N(l)$, or the number of time windows, $N(n)$, in assortment T. This approach allows us to analyze both relationships separately and independently.

3.3.3. Delivery Model

Based on the demand per time window $N(T)$ and assortment T, the delivery model evaluates the ensuing delivery costs and the required vehicle capacity. We first model the case of *consecutive* time windows, and then expand the model to deal with *overlapping* time windows.

Consecutive Time Windows

We consider time window assortments $T \in \mathcal{T} = \{(l, n, o) \mid l \in \mathbb{R}^+, n \in \mathbb{N}^+, o = \text{con.}\}$, representing n consecutive time windows of length l, and resulting in a delivery shift of length $L(T) := nl$. Following the continuous approximation approach, we distinguish two components of a vehicle tour – the *stem distance* to and from the delivery area and the *circular distance*, i.e., the total distance between customers (see C. F. Daganzo, 1984). In our case, the circular distance stretches across several consecutive time windows. Figure 3.4a illustrates the different elements of a tour.



(a) Tour components **(b)** Customer densities

Figure 3.4.: Graphical representation of vehicle tours over consecutive time windows

To assess the total travel costs, we first approximate the travel time between customers within a time window. We express the average distance between customers as a function of customer density, $\frac{k}{\sqrt{\delta}}$. For a time window demand $N(T)$ and a delivery region of size R, the customer density in a time window is $\delta(T) := \frac{N(T)}{R}$ (Figure 3.4b). The inter-customer distance then becomes $k\sqrt{\frac{R}{N(T)}}$ which follows a Beardwood-Halton-Hammersley routing distance that captures marginal efficiency gains for an increasing customer density (Beardwood et al., 1959). Second, we express the stem distance as twice the average distance between the center of the delivery region and any point within the region. Under the Euclidean metric, this results in a stem distance of $\frac{4}{3}\sqrt{\frac{R}{\pi}}$ per tour.

The time window length limits the number of deliveries a single vehicle can make in that time window. To model this effect, we introduce the concept of *workload* per time window, which denotes the total time required to satisfy the demand in each time window. This total time combines the service time needed to perform a delivery, e.g., parking, unloading etc., and the travel time between customers. Note that the stem distance can be traversed outside of the delivery shift and is therefore not affected by the time windows. Assuming a constant vehicle speed α and a constant service time per order τ, the workload per time window becomes $w(T) := N(T) \cdot \left[\tau + \alpha k\sqrt{\frac{R}{N(T)}}\right] = \tau N(T) + \alpha k\sqrt{RN(T)}$, which corresponds to a workload per customer of $w^c(T) := \tau + \alpha k\sqrt{\frac{R}{N(T)}}$.

Operationally, the required fleet size depends on both time window length and vehicle capacity, as these factors limit the number of customers a vehicle can serve on a single tour. Because we focus on strategic time window assortment design, assortment decisions precede fleet and capacity planning. Consequently, we consider time constraints, rather than vehicle capacity, as the primary driver of fleet size when evaluating a time window assortment. Given a workload per time window $w(T)$, the *number of vehicle tours* required to satisfy the time window demand is $v(T) := \frac{w(T)}{l}$.

The vehicle capacity is treated as an endogenous outcome of our model, determining the vehicle size required to operate the number of tours dictated by the time window constraints. Consequently, the time window assortment drives the required vehicle capacity. Formally, given time window assortment $T \in \mathcal{T}$, the number of orders that a delivery vehicle needs to carry is

$$Q(T) := \frac{nN(T)}{v(T)} = \frac{L(T)}{w^c(T)}. \tag{3.1}$$

We now have all the elements to approximate the total delivery costs. The *fixed cost* of operating a vehicle tour includes the vehicle-related cost f and the stem cost per tour. With a cost per minute of c, the fixed cost per tour becomes $F := f + c\alpha\frac{4}{3}\sqrt{\frac{R}{\pi}}$. For the *variable cost*, we use the same time-based cost factor c to estimate the cost per tour. This cost accounts for both travel and service activities during the shift and depends on the delivery shift length L. Consequently, the delivery cost per tour is $C^t(T) := cL(T) + F$, and the total delivery costs are given by $C(T) := C^t(T)v(T)$. In summary, our delivery model leads to the following expression for total profit given assortment $T \in \mathcal{T}$:

$$P(T) := S(T) - C(T) = \Big(r - cw^c(T)\Big)nN(T) - Fv(T). \tag{3.2}$$

Overlapping Time Windows

The route approximation outlined above evaluates consecutive time windows. We now introduce a demand allocation procedure that allows our route approximation also to handle assortments with overlapping time windows. In line with previous research (N. Agatz et al., 2011; C. F. Daganzo, 1987), we optimally allocate the total demand for the overlapping assortment to a corresponding set of consecutive intervals such that the total delivery costs are minimized. Figure 3.5 provides a graphical representation of the overlapping time window assortment and the corresponding consecutive delivery intervals.

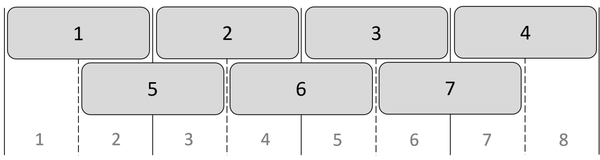


Figure 3.5.: Consecutive delivery intervals for an assortment of $n = 7$ overlapping time windows

Formally, we consider overlapping time window assortments denoted by $T \in \widetilde{\mathcal{T}} = \{(l, n, o) \mid l \in \mathbb{R}^+, n = 2a + 1, a \in \mathbb{N}^+, o = \text{ovl.}\}$ that consist of n overlapping time windows of length l. These time windows are divided into $(n + 1)$ consecutive delivery intervals of length $\frac{l}{2}$, which results in a delivery shift length of $\widetilde{L}(T) := \frac{n+1}{2}l$.

Lemma 3.1. *An even distribution of demand across consecutive delivery intervals minimizes the total delivery costs.*

Following Lemma 3.1, a demand of $\frac{n}{n+1}N(T)$ is allocated to each consecutive delivery interval, leading to a workload per customer of $\widetilde{w}^c(T) := \tau + \alpha k\sqrt{\frac{n+1}{n}\frac{R}{N(T)}}$, a workload per interval of $\widetilde{w}(T) := \tau\frac{n}{n+1}N(T) + \alpha k\sqrt{R\frac{n}{n+1}N(T)}$, and $\widetilde{v}(T) := \frac{2\widetilde{w}(T)}{l}$ vehicle tours. Given assortment $T \in \widetilde{\mathcal{T}}$, the number of orders that a delivery vehicle needs to carry is

$$\widetilde{Q}(T) := \frac{nN(T)}{\widetilde{v}(T)} = \frac{\widetilde{L}(T)}{\widetilde{w}^c(T)}. \tag{3.3}$$

Furthermore, given a delivery cost per tour of $\widetilde{C}^t(T) := c\widetilde{L}(T) + F$ for assortment $T \in \widetilde{\mathcal{T}}$, the total delivery cost is given by $\widetilde{C}(T) := \widetilde{C}^t(T)\widetilde{v}(T)$, leading to a total profit of

$$\widetilde{P}(T) := S(T) - \widetilde{C}(T) = \Big(r - c\widetilde{w}^c(T)\Big)nN(T) - F\widetilde{v}(T). \tag{3.4}$$

3.4. Profitability Drivers and Operational Constraints

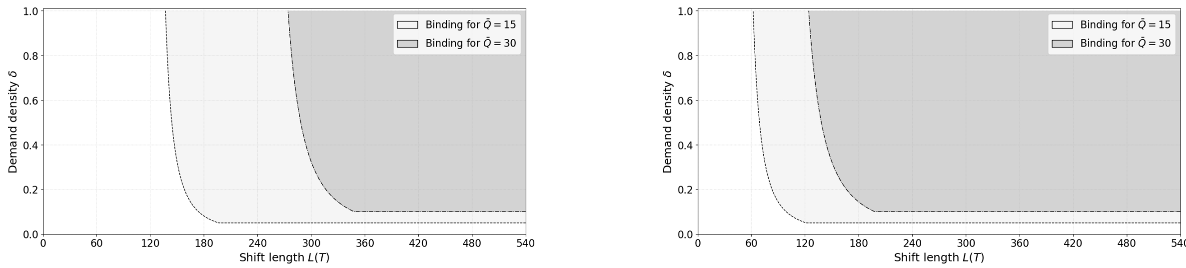
In this section, we assess the financial and operational performance of different *consecutive* time window assortments under varying demand levels, which we treat as an

Table 3.1.: Summary of notation used in Chapter III

General Model Input	
α	Driving speed [min/km]
c	Variable cost per time unit [EUR/min]
f	Vehicle-related fixed cost per tour [EUR]
F	Fixed cost per tour [EUR]
k	Road network factor
r	Gross margin per order [EUR]
R	Surface of the delivery region [km^2]
τ	Delivery service time per order [min]
Time Window Assortment Input	
l	Time window length [min]
n	Number of time windows
o	Consecutive/ overlapping flag
$T \in \mathcal{T} \cup \tilde{\mathcal{T}}$	Time window assortment
Performance Output	
C^t	Delivery cost per tour [EUR]
δ	Demand density per time window [orders/km^2]
L	Delivery shift length [min]
N	Demand per time window [orders]
v	Number of vehicle tours
w	Workload per time window [min]
w^c	Workload per customer [min]

exogenous parameter. Formally, we consider assortments $T \in \mathcal{T}$ with a demand of $N \in \mathbb{R}^+$ orders per time window. Additionally, we introduce a maximum capacity threshold $\bar{Q} \in \mathbb{R}^+$, which represents a practical upper bound on vehicle size, reflecting real-world constraints such as vehicle availability or regulatory limits. This threshold ensures that only time window assortments resulting in feasible vehicle capacities are considered, i.e., those for which $Q(T, N) \leq \bar{Q}$.

First, we analyze the parameter regions where the required vehicle capacity approaches the practical *feasibility limits* defined by $\bar{Q}$. In low-density areas, where travel times between stops are relatively long, vehicle capacity constraints often become negligible, as time window restrictions tend to dominate route feasibility. This explains why early research in grocery delivery primarily focused on time window constraints. However, as demand densities increase, vehicle capacity becomes increasingly binding. Online



(a) $\tau = 8$ (b) $\tau = 3$

Figure 3.6.: Parameter regions in the (L, δ)-space where capacity $\bar{Q} \in \{15, 30\}$ becomes binding for varying time window lengths $l \in (0, 180]$ with $n = 3$ time windows and varying demand levels $N \in (0, 100]$ within a region of size $R = 100$ ($\alpha = 2.0$, $k = 0.57$)

grocers must then either deploy larger vehicles (van Brouwershaven, 2020) or shorten delivery shifts to accommodate the growing demand.

To anticipate this effect, we derive analytical conditions that ensure that the required vehicle capacity $Q(T, N)$ for a time window assortment T, facing demand density $\delta = \frac{N}{R}$, remains *feasible* within the practical upper bound on vehicle capacity, $\bar{Q}$. These conditions help align the time window assortment with practical fleet constraints for varying demand densities, thereby helping service providers anticipate fleet requirements effectively.

Lemma 3.2. *Consider Equation* (3.1) *for a given demand density* $\delta = \frac{N}{R} > 0$ *and a maximum vehicle capacity* $\bar{Q} > 0$*. For time window assortments* $T \in \mathcal{T}$*, the maximum vehicle capacity is binding if and only if*

$$\frac{\bar{Q}}{n} < \min\left\{\delta R, \frac{l}{\tau + \frac{\alpha k}{\sqrt{\delta}}}\right\}$$

There are three factors that limit the number of orders served on a tour: the demand level, the physical vehicle capacity, and the available time. The condition above states that, intuitively, the vehicle capacity is binding if it is more constraining than the other two factors. For example, in low-density settings, i.e., when $\delta R \leq \frac{\bar{Q}}{n}$, vehicle capacity is not a restriction. However, once the density exceeds this threshold, i.e., $\delta R > \frac{\bar{Q}}{n}$, vehicle capacity may become a limiting factor, particularly when time windows are long.

Figure 3.6 shows regions in the (L, δ)-space where vehicle capacity is binding, based on varying time window lengths $l \in (0, 180]$ (expressed as shift length $L(T) = nl$, with

$n = 3$), varying demand $N \in (0, 100]$ (expressed as density $\delta = \frac{N}{R}$, with $R = 100$), and two values of the capacity bound $\bar{Q}$. The graphs illustrate how anticipating physical capacity constraints limits the range of feasible time window lengths.

For all subsequent analyses, we further introduce the notion of an *efficient* time window assortment. Specifically, a time window assortment T is considered efficient under demand N if it fills at least one complete vehicle tour, i.e., if $v(T, N) \geq 1$.

Corollary 3.1. *For any assortment $T \in \mathcal{T}$ that is efficient under demand density $\delta = \frac{N}{R} > 0$, it holds that:*

$$\delta R \geq \frac{l}{\tau + \frac{\alpha k}{\sqrt{\delta}}}$$

Having discussed the limits of feasibility, we now consider the *profitability* of time window assortments. We derive analytical expressions that allow us to identify key drivers of profitability, defined by the condition $P(T, N) \geq 0$.

Theorem 3.1. *Consider Equations* (3.1) *and* (3.2) *and a maximum vehicle capacity $\bar{Q} > 0$, and let the net margin per order be positive, $r > c\tau$. Furthermore, let $N^*(n) = \frac{2(r-c\tau)F+n(c\alpha k)^2 R}{2n(r-c\tau)^2} + \sqrt{\left(\frac{2(r-c\tau)F+n(c\alpha k)^2 R}{2n(r-c\tau)^2}\right)^2 - \left(\frac{F}{n(r-c\tau)}\right)^2}$. Then, the delivery shift length $L(T)$ defines three performance ranges of time window assortments $T \in \mathcal{T}$:*

a) [Unprofitable region] $L(T) \leq \frac{F\tau}{r-c\tau}$: It is impossible to achieve profitability regardless of demand density.

b) [Multi-tour profit region] $L(T) \in \left(\frac{F\tau}{r-c\tau}, \frac{Fw^c(N^*(n))}{r-cw^c(N^*(n))}\right)$: The minimum demand density $\delta^P(T) := \frac{N^P(T)}{R}$ for which assortment T is profitable is $\left(\frac{C^t(T)\alpha k}{rL(T)-C^t(T)\tau}\right)^2$. For any density $\delta \geq \delta^P(T)$ and any shift length $L(T)$ within the specified range, the required vehicle capacity satisfies $Q(T, \delta) \geq \frac{L(T)}{\tau+\frac{\alpha k}{\sqrt{\delta^P(T)}}}$.

c) [Single-tour profit region] $L(T) \geq \frac{Fw^c(N^*(n))}{r-cw^c(N^*(n))}$: The minimum demand density $\delta^v(l) := \frac{N^v(l)}{R}$ to profitably operate assortment T with a single vehicle tour is $\frac{2l\tau+(\alpha k)^2 R-\sqrt{(2l\tau+(\alpha k)^2 R)^2-(2l\tau)^2}}{2R\tau^2}$. For any density $\delta \geq \delta^v(l)$ and any shift length $L(T)$ within the specified range, the required vehicle capacity satisfies $Q(T, \delta) \geq n\delta^v(l)R$.

Theorem 3.1 shows that the shift length is an important factor affecting the profitability of the delivery system. More specifically, it is the shift length in relation to the ratio of fixed costs to variable contribution margins that determines whether profitability can be achieved or not. Theorem 3.1a) states a lower bound on the shift length for any positive profit. This lower bound equals the service time needed for the minimum number of orders required for covering the fixed costs per tour. For shifts longer than this lower bound, demand is another relevant constraint. Theorem 3.1b) shows that there exists a break-even demand density above which profit is positive and increasing in demand.

If the shift length is increased further, it will eventually equal the total time required to serve the corresponding break-even demand, expressed as $nN^*(n)w^c\left(N^*(n)\right)$, where $N^*(n)$ denotes the break-even demand for a single vehicle tour. At this point, the following relationship holds: $nN^*(n) = \frac{F}{r-cw^c(N^*(n))}$. Thus, for shift lengths beyond the upper bound stated in Theorem 3.1c), we need to ensure a minimum demand density to run a single vehicle tour and not create unnecessary slack time during the delivery shift. As this minimum density exceeds the density level that would be needed to break even, an efficient shift length falls within the boundaries specified in Theorem 3.1b).

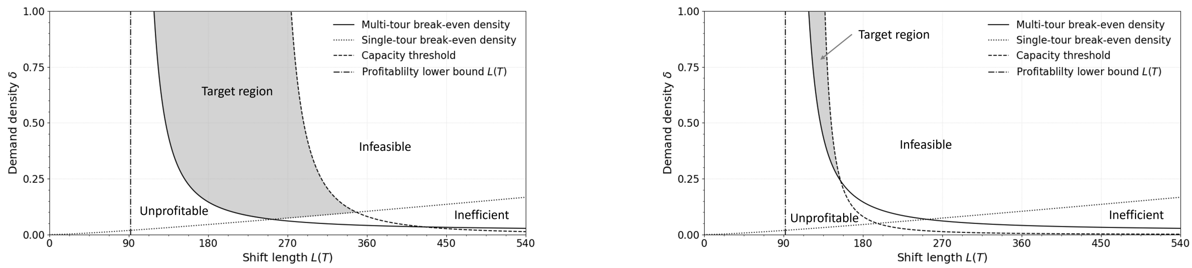
In addition, to determine whether assortment T leads to a feasible vehicle capacity requirement under demand N, i.e., whether $Q(T,N) \leq \bar{Q}$, we apply the following equivalent transformation:

Corollary 3.2. *Consider Equation* (3.1) *and a maximum vehicle capacity* $\bar{Q} > 0$. *For any assortment* $T \in \mathcal{T}$ *and any demand density* $\delta = \frac{N}{R} > 0$, *the following equivalence holds:*

$$Q(T,N) \leq \bar{Q} \quad \Leftrightarrow \quad \delta \leq \begin{cases} \left(\frac{\bar{Q}\alpha k\sqrt{R}}{L(T)-\bar{Q}\tau}\right)^2, & L(T) > \bar{Q}\tau \\ \infty, & L(T) \leq \bar{Q}\tau \end{cases}$$

Figure 3.7 shows regions in the (L,δ)-space where we highlight the target performance region which we distinguish from unprofitable delivery, $P(T,N) < 0$, inefficient delivery, $v(T,N) < 1$, and deliveries that are infeasible with respect to the maximum vehicle capacity threshold, $Q(T,N) > \bar{Q}$. The plots are based on varying time window lengths $l \in (0,180]$ (expressed as shift length $L(T) = nl$, with $n = 3$), varying demand $N \in (0,100]$ (expressed as density $\delta = \frac{N}{R}$, with $R = 100$), and two values of the capacity bound $\bar{Q} \in \{15,30\}$.

The graphs illustrate that the break-even density decreases as the shift length increases. That is, at higher densities, we can become profitable even with shorter shifts and shorter time windows. We also observe that once we break even, extending the shift



(a) $\bar{Q} = 30$ **(b)** $\bar{Q} = 15$

Figure 3.7.: Performance regions distinguishing target performance from unprofitable ($P(T, N) < 0$), inefficient ($v(T, N) < 1$), and infeasible ($Q(T, N) > \bar{Q}$) outcomes for varying time window lengths $l \in (0, 180]$ with $n = 3$ time windows and varying demand levels $N \in (0, 100]$ within a region of size $R = 100$ ($r = 9.0$, $c = 0.5$, $f = 50.0$, $\tau = 8.0$, $\alpha = 2.0$, $k = 0.57$)

length further will eventually result in either reaching the maximum physical vehicle capacity in high-density scenarios or facing inefficiencies in tour utilization in low-density scenarios. Figure 3.7b shows how a more constrained vehicle size results in a more narrow target performance region.

3.5. Adjusting the Time Window Length

Having identified the key factors that drive time window assortment profitability under the operational constraints discussed, we now turn to incremental assortment design *decisions*. To this end, we treat demand as endogenous to the time window assortment. Throughout all subsequent analyses, we ensure that the time window length in assortment T allows for at least one complete vehicle tour, $l \leq w(T)$ for $T \in \mathcal{T}$ and $\frac{l}{2} \leq \widetilde{w}(T)$ for $T \in \widetilde{\mathcal{T}}$, and that the required vehicle capacities remain within the feasible limits, $Q(T) \leq \bar{Q}$ for $T \in \mathcal{T}$ and $\widetilde{Q}(T) \leq \bar{Q}$ for $T \in \widetilde{\mathcal{T}}$.

In this section, we consider how to select the time window length for a fixed number of *consecutive* time windows, based on total profit. To do so, we introduce endogenous demand $N(l)$, where the demand per time window depends on the time window length l offered in assortment $T \in \mathcal{T}$. We focus on two extremes in terms of customer choice behavior: perfect demand substitution (Section 3.5.1) and no demand substitution (Section 3.5.2).

3.5.1. Perfect Demand Substitution

In this section, we analyze time window length decisions when customers are insensitive to time window length – that is, they perceive time windows of different lengths as perfect substitutes. Therefore, we express demand per time window independent of the time window length as $N(l) = \hat{N}$.

We begin our analysis by identifying the main trade-offs underlying time window length decisions. The following result illustrates how the time window length impacts the various individual (cost) components of a delivery system.

Lemma 3.3. *Consider Equations* (3.1) *and* (3.2) *for a given number of* $n \geq 1$ *consecutive time windows and demand per time window* $N(l)$ *that is independent of the time window length,* $\frac{\partial N}{\partial l}(l) = 0$*. For time window lengths* $l > 0$ *and demand* $N(l) > 0$*, we get:*

a) Workload $w(l)$ *and workload per customer* $w^c(l)$ *are constant in the time window length* l*.*

b) Number of tours and associated costs decrease in the time window length l*:* $\frac{\partial v}{\partial l}(l) < 0$.

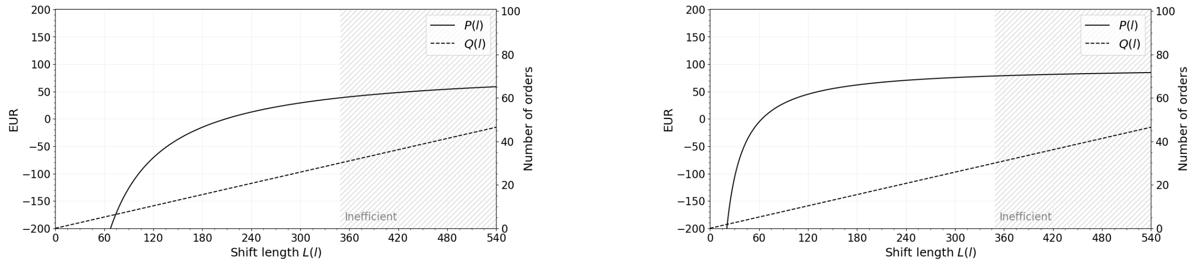
c) Profit increases in the time window length l*:* $\frac{\partial P}{\partial l}(l) > 0$.

d) Required vehicle capacity increases in the time window length l*:* $\frac{\partial Q}{\partial l}(l) > 0$.

Enlarging time windows increases profits by reducing fleet size and fixed costs, but this requires larger vehicles to meet the increased demand per tour. The system requires fewer vehicle tours, but larger vehicles for each tour. Conversely, decreasing the time window length leads to the opposite effects: more vehicle tours are required, but the required vehicle capacity decreases as a result of smaller demand per tour.

It follows that, under perfect substitution, wider time windows consistently lead to higher profits, while narrower time windows reduce profits. This result is formally stated in Proposition 3.1.

Proposition 3.1. *Consider Equations* (3.1) *and* (3.2) *for a given number of* $n \geq 1$ *consecutive time windows and demand per time window* $N(l) = \hat{N} > 0$*. For time windows of length* $l < w(l)$ *and sufficiently large maximum vehicle capacity* $Q(l) \leq \bar{Q}$*, profit and the required vehicle capacity both increase in the time window length.*



(a) Profit and required capacity with $f = 50$

(b) Profit and required capacity with $f = 10$

Figure 3.8.: Profitability of length adjustments for varying time window lengths $l \in (0, 180]$ with $n = 3$ consecutive time windows ($r = 9.0$, $c = 0.5$, $\tau = 8.0$, $\alpha = 2.0$, $k = 0.57$, $R = 100$, $\hat{N} = 10$)

Figure 3.8 provides two numerical examples showing the total profit $P(l)$ and required vehicle capacity $Q(l)$ for a given demand density of $\frac{\hat{N}}{R} = 0.1$ per time window in a delivery region of size $R = 100$, and $n = 3$ consecutive time windows of varying length $l \in (0, 180]$ (expressed as shift length $L(l) = nl$). Additionally, we mark the parameter region where the time window length exceeds the workload, $l > w(l)$, indicating inefficiencies due to underutilized time. We conclude that under perfect substitution, service providers should offer the longest time window that still ensures efficient tours, i.e., $l = w(l)$.

3.5.2. No Demand Substitution

Complementing the previous section, we now analyze time window length decisions in a setting where customers are sensitive to time window length. These customers have a maximum acceptable time window length and reject longer alternatives. Specifically, we capture time window length on a continuous scale and approximate equal-sized discrete customer segments using a continuous spectrum. This approach allows us to model the time window demand as a linear function of the time window length, expressed as $N(l) = \bar{N} - \gamma l$, where $\bar{N}$ represents the demand potential that can be captured per time window when the service provider offers the shortest, most attractive time window length (i.e., an exact delivery time, $l = 0$). The parameter γ is a time sensitivity factor that quantifies how rapidly demand declines as the length of the time window increases. To ensure positive demand for each time window, the time window length must not exceed the maximum acceptable length across all customer segments, i.e., $l < \frac{\bar{N}}{\gamma}$.

We begin our analysis again by identifying the main trade-offs underlying time window length decisions. The following results illustrate how the time window length impacts the various individual (cost) components of a delivery system.

Lemma 3.4. *Consider Equations* (3.1) *and* (3.2) *for a given number of* $n \geq 1$ *consecutive time windows and demand per time window* $N(l)$ *that is decreasing in the time window length,* $\frac{\partial N}{\partial l}(l) < 0$. *For time window lengths* $l > 0$ *and demand* $N(l) > 0$, *we get:*

a) Unit service time is constant but inter-customer travel time increases in the time window length l. *Thus, the workload per customer increases:* $\frac{\partial w^c}{\partial l}(l) > 0$. *As a result, the contribution margin per customer,* $r - cw^c(l)$, *decreases for longer time windows.*

b) Both service time and travel time between customers decrease in the time window length l. *Thus, the workload decreases:* $\frac{\partial w}{\partial l}(l) < 0$. *As a result, the associated variable costs decrease for longer time windows.*

c) Number of tours decreases in the time window length l: $\frac{\partial v}{\partial l}(l) < 0$. *As a result, the required fleet size and the associated fixed costs decrease for longer time windows.*

d) Profit is not monotonic in the time window length l:
$\frac{\partial P}{\partial l}(l) = \left(n(r - c\tau) - \frac{1}{2}nc\alpha k\sqrt{\frac{R}{N(l)}}\right)\frac{\partial N}{\partial l}(l) - F\frac{\partial v}{\partial l}(l)$.

e) Required vehicle capacity is not monotonic in the time window length l:
$\frac{\partial Q}{\partial l}(l) = \frac{n}{w^c(l)}\left(1 - \frac{l}{w^c(l)}\frac{\partial w^c}{\partial l}(l)\right)$.

From Lemma 3.4, we conclude that the appropriate time window length results from trading off benefits of additional demand against the cost of additional vehicle tours. Smaller time windows attract more demand, resulting in higher sales and more efficient travel between customers, but they also correspond to a higher total workload and less time available for delivery and, therefore, more vehicle tours. Formally, shortening time windows increases profits if and only if the gain from additional demand exceeds the cost of making additional vehicle tours. Furthermore, shortening time windows decreases the required vehicle size if and only if the decrease in workload per customer is small enough. Conversely, increasing the time window length has the opposite effect.

Building on these insights, we derive analytical expressions to identify the conditions for adjusting the time window length to increase the total profit.

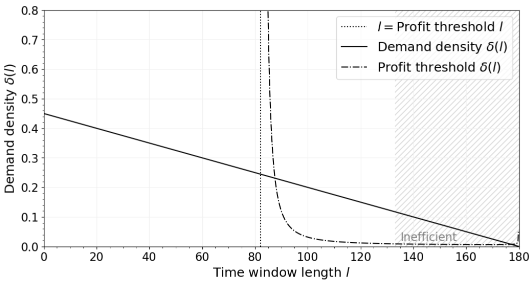
Proposition 3.2. *Consider Equations* (3.1) *and* (3.2) *for a given number of* $n \geq 1$ *consecutive time windows and demand* $N(l) = \bar{N} - \gamma l$ *with a demand potential of* $\bar{N} > 0$ *per time window and a time sensitivity factor* $\gamma > 0$. *Let the net margin per order be positive,* $r > c\tau$. *For time windows of length* $l < w(l)$ *and sufficiently large maximum vehicle capacity* $Q(l) \leq \bar{Q}$, *increasing the time window length* l *increases profit if and only if*

$$l \leq \frac{\sqrt{\mathcal{D}(l) + \mathcal{W}(l)} - \sqrt{\mathcal{D}(l)}}{4n(r - c\tau)} \qquad \text{or} \qquad \delta(l) < \left(\frac{\gamma n c \alpha k}{2\mathcal{V}(l)}\right)^2,$$

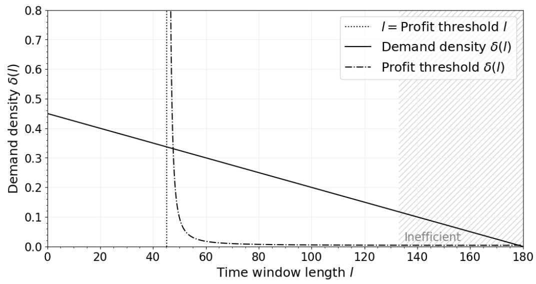
with demand density $\delta(l) = \frac{N(l)}{R}$ *as well as* $\mathcal{D}(l) := \frac{(F\alpha k)^2}{\delta(l)}$, $\mathcal{W}(l) := 16n(r - c\tau)\frac{F}{\gamma}\bar{N}w^c(l)$, *and* $\mathcal{V}(l) := \gamma n(r - c\tau) + F\frac{\partial v}{\partial l}(l)$. *For* $l \leq \frac{2}{3}\frac{\bar{N}}{\gamma}$, *the required vehicle capacity is increasing,* $\frac{\partial Q}{\partial l}(l) > 0$.

Proposition 3.2 reveals that two factors determine whether increasing the window length of a time window assortment is beneficial from a profit perspective or not, namely (i) the current time window length and (ii) demand density. If either of these factors is sufficiently low profit increases by increasing l, otherwise it does not. The result follows from the trade-off identified in Lemma 3.4d). The relevance of l is intuitive. The role of $\delta(l)$, however, is less straightforward. Lower customer densities are associated with higher workloads and lower contribution margins per customer, as shown in Lemma 3.4a). We therefore hypothesize that, in this case, the operational benefits of additional delivery time outweigh the comparatively modest financial losses resulting from reduced demand due to longer time windows – two opposing effects that are also reflected in Lemma 3.4d). Additionally, we identify a sufficient condition for the required vehicle capacity to increase in the length of the time windows. This insight helps determine whether the vehicles required to accommodate the additional demand resulting from extended time windows remain within the capacity limit $\bar{Q}$. For shortening the time windows, the reverse applies.

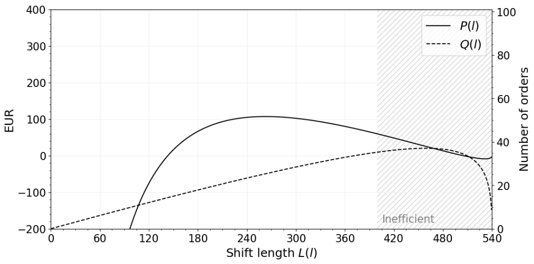
Figure 3.9 illustrates numerical examples of the profit conditions stated in Proposition 3.2 (Figures 3.9a and 3.9b), as well as the trajectories of total profit $P(l)$ and required vehicle capacity $Q(l)$ (Figures 3.9c and 3.9d) for a given demand density potential of $\frac{\bar{N}}{R} = 0.45$ in a delivery region of size $R = 100$, a time sensitivity of $\gamma = 0.25$, and $n = 3$ consecutive time windows of varying length $l \in (0, \frac{\bar{N}}{\gamma})$. We consider $l < \frac{\bar{N}}{\gamma}$ to ensure positive demand, and we highlight the parameter region where the time window length exceeds the workload, $l > w(l)$, indicating inefficiencies due to underutilized time.



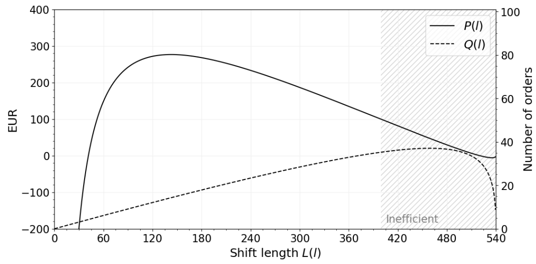
(a) Trade-off with $f = 50$



(b) Trade-off with $f = 10$



(c) Profit and required capacity with $f = 50$



(d) Profit and required capacity with $f = 10$

Figure 3.9.: Profitability of length adjustments for varying time window lengths $l \in (0, \frac{\bar{N}}{\gamma})$ with $n = 3$ consecutive time windows ($r = 9.0$, $c = 0.5$, $\tau = 8.0$, $\alpha = 2.0$, $k = 0.57$, $R = 100$, $\bar{N} = 45$, $\gamma = 0.25$)

The graphs show that the optimal time window length increases with higher fixed costs f, where the additional time needed to reduce the number of vehicle tours becomes more valuable than the revenue loss when capacity is tight. Moreover, Figures 3.9a and 3.9b suggest that the implicit condition on the time window length l, stated in Proposition 3.2, serves as a good indicator for when to increase or decrease the time window length, as the optimum occurs approximately at $l = \frac{\sqrt{\mathcal{D}(l)+\mathcal{W}(l)}-\sqrt{\mathcal{D}(l)}}{4n(r-c\tau)}$.

3.6. Adjusting the Number of Time Windows

In this section, we address the decision of selecting the number of consecutive and overlapping time windows of a fixed length l to offer in a time window assortment. We construct *consecutive* assortments by sequencing $\hat{n}$ consecutive time windows of equal length, and analyze how incremental changes in $\hat{n}$ impact profit. In the case of *overlapping* assortments, we evaluate the decision to add $\hat{n} - 1$ overlapping time windows to consecutive assortments $(l, \hat{n}, \text{con.}) \in \mathcal{T}$. This results in new assortments $(l, 2\hat{n}-1, \text{ovl.}) \in \widetilde{\mathcal{T}}$ with equal shift length of $L(\hat{n}) = \widetilde{L}(2\hat{n}-1) = \hat{n}l$.

Furthermore, we introduce endogenous demand $N(n)$, where the demand per time window depends on the number of time windows n offered within any assortment T. As in the previous section, we focus on two extremes in terms of choice behavior: perfect demand substitution (Section 3.6.1) and no demand substitution (Section 3.6.2).

3.6.1. Perfect Demand Substitution

We analyze the impact of adjusting the number of time windows in assortment T, assuming that all time windows are perfect substitutes. This means that while customers may have individual preferences, they are willing to substitute among the available time windows. Consequently, the total demand volume, denoted by $\bar{N}$, remains constant, regardless of the number of time windows offered. Under the homogeneity assumption (Assumption 3.2), customers are evenly distributed across the n time windows. As a result, the demand per time window is $N(n) = \frac{\bar{N}}{n}$.

In the following, we derive analytical expressions to identify the conditions under which incrementally adjusting the number of time windows increases total profit.

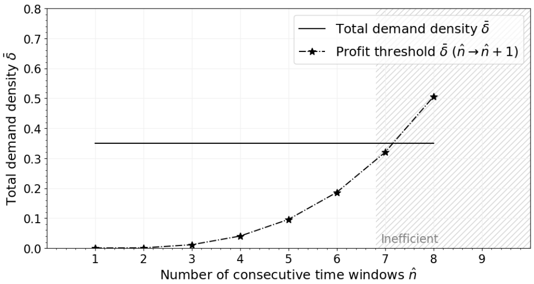
Proposition 3.3. *Consider Equations* (3.1) *to* (3.4) *for a given time window length* $l > 0$ *and demand* $N(n) = \frac{\bar{N}}{n}$ *with a total demand volume* $\bar{N} \geq 2N^v(l)$.

1. *For* $1 \leq \hat{n} \leq \frac{\bar{N}}{N^v(l)} - 1$ *consecutive time windows and sufficiently large maximum vehicle capacity* $Q(\hat{n}) \leq \bar{Q}$*, adding a consecutive time window increases profits if and only if*

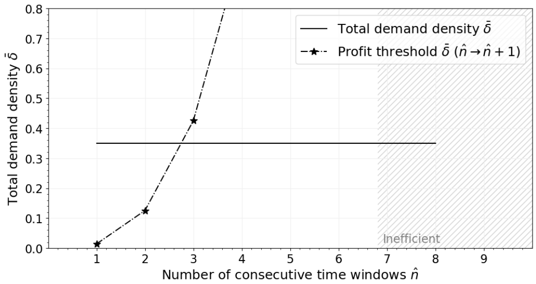
$$\bar{\delta} > \left(\frac{\alpha k \sqrt{(\hat{n}+1)\hat{n}}}{\tau(\sqrt{\hat{n}+1} + \sqrt{\hat{n}})} \right)^2 \left(\frac{lc\sqrt{(\hat{n}+1)\hat{n}}}{F} - 1 \right)^2,$$

   with total demand density $\bar{\delta} = \frac{\bar{N}}{R}$*. Furthermore,* $Q(n+1) = \frac{v(n)}{v(n+1)} Q(n)$*, which implies that the required vehicle capacity increases.*

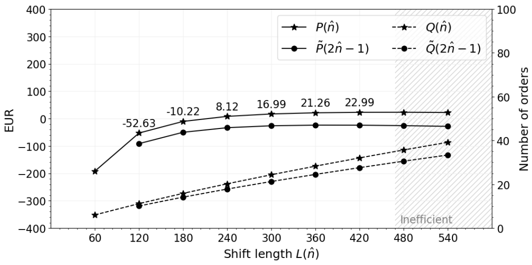
2. *For* $2 \leq \hat{n} \leq \frac{\bar{N}}{N^v(l)}$ *consecutive time windows and sufficiently large maximum vehicle capacity* $Q(\hat{n}) \leq \bar{Q}$*, adding* $\hat{n}-1$ *overlapping time windows always decreases profits. However, the required vehicle capacity also decreases.*



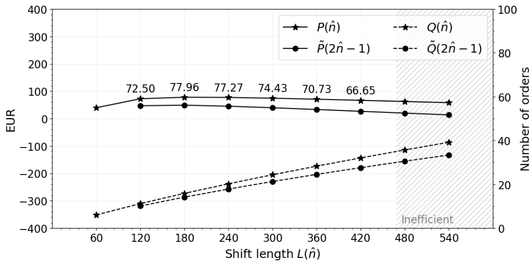
(a) Trade-off with $f = 50$



(b) Trade-off with $f = 10$



(c) Profit and required capacity with $f = 50$



(d) Profit and required capacity with $f = 10$

Figure 3.10.: Profitability of adjusting the number of time windows for varying number of consecutive time windows $\hat{n} \in \{1, \ldots, 9\}$ each of length $l = 60$ ($r = 9.0$, $c = 0.5$, $\tau = 8.0$, $\alpha = 2.0$, $k = 0.57$, $R = 100$, $\bar{N} = 35$)

Interestingly, Proposition 3.3.1 shows that even when total demand remains unaffected by the number of time window options, adding consecutive time windows can be beneficial when demand density is sufficiently high. In this case, the savings in fixed delivery costs from reducing the number of delivery tours outweighs the increase in variable travel costs. This makes it beneficial to spread demand across fewer but longer vehicle tours. Conversely, at lower densities, reducing the number of consecutive time windows may be preferable. The key trade-off is between spreading demand to minimize required vehicle tours and clustering demand to reduce average travel distance per delivery. Balancing these opposing effects is crucial for determining the appropriate number of consecutive time windows.

However, this is not the case for overlapping time windows. Proposition 3.3.2 shows that when demand is unaffected by the number of time windows, adding overlapping time windows does not increase profits. This is intuitive, since overlapping time windows do not help to spread demand but rather add more time constraints within the same time frame. This suggests that overlapping time windows may actually hurt profitability.

Figure 3.10 illustrates numerical examples of the profit condition stated in Proposition 3.3.1 (Figures 3.10a and 3.10b), as well as the trajectories of total profit and required vehicle capacity (Figures 3.10c and 3.10d) for a given total demand density of $\frac{\bar{N}}{R} = 0.35$ in a delivery region of size $R = 100$, and a varying number of consecutive time windows $\hat{n} \in \{1, \ldots, 9\}$ each of length $l = 60$ minutes. Both Figures 3.10c and 3.10d confirm that profit and required vehicle capacity decrease as overlapping time windows are added. For consecutive time windows, Figure 3.10c shows that when fixed costs f are relatively high, the total demand density remains sufficiently large for profit to continue increasing as demand is spread, up until the workload per time window falls below the time window length.

Conversely, for relatively low fixed costs, Figure 3.10d highlights a trade-off between different cost elements: the profit function initially increases before declining, given a constant gross margin of $r\bar{N}$. This trade-off is further illustrated in Figure 3.10b, where at $\hat{n} = 3$ consecutive time windows, adding another time window reduces profits because the total demand density is too low for the marginal decrease in fixed costs to outweigh the marginal increase in variable costs. In summary, the optimal number of time windows is the smallest value at which the condition in Proposition 3.3.1 does no longer hold.

3.6.2. No Demand Substitution

Next, we analyze adjusting the number of time windows in assortment T, assuming customers do not substitute. This means that distinct customer segments of equal size have a preferred time window and will not accept any alternative. As a result, the total demand $nN(n)$ increases linearly with the number of time windows. Formally, we express the time window demand as $N(n) = \hat{N}$, where $\hat{N}$ represents the size of the customer segments.

In the following, we derive analytical expressions to identify the conditions under which incrementally adjusting the number of time windows increases total profit.

Proposition 3.4. *Consider Equations* (3.1) *to* (3.4) *for a given time window length* $l > 0$ *and demand* $N(n) = \hat{N} \geq N^v(l)$. *Let the net margin per order be positive,* $r > c\tau$.

1. *For $\hat{n} \geq 1$ consecutive time windows and sufficiently large maximum vehicle capacity $Q(\hat{n}) \leq \bar{Q}$, adding a consecutive time window increases profits for any $\hat{n}$ if and only if*

$$\hat{\delta} > \left(\frac{c\alpha k}{r - c\tau}\right)^2 \qquad \Leftrightarrow \qquad r - c\tau > \frac{c\alpha k}{\sqrt{\hat{\delta}}},$$

   with demand density per time window $\hat{\delta} = \frac{\hat{N}}{R}$. Furthermore, $Q(n+1) = Q(n) + Q(1)$, which implies that the required vehicle capacity increases.

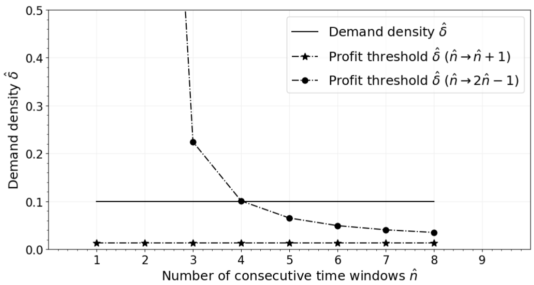
2. *For $\hat{n} \geq 2$ consecutive time windows, shift length $L(\hat{n}) > \frac{F\tau}{r - c\tau}$, and sufficiently large maximum vehicle capacity $Q(\hat{n}) \leq \bar{Q}$, adding $\hat{n} - 1$ overlapping time windows increases profits if and only if*

$$\hat{\delta} > \left(\frac{\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}}{\hat{n} - 1}\right)^2 \delta^P(\hat{n}),$$

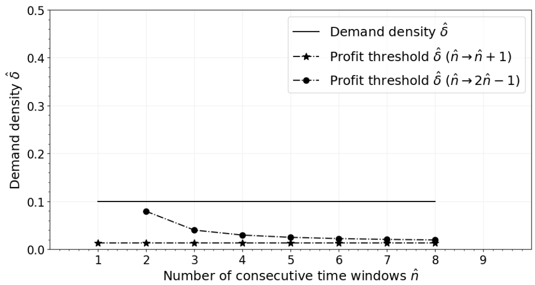
   with demand density per time window $\hat{\delta} = \frac{\hat{N}}{R}$. It holds that $\left(\frac{\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}}{\hat{n}-1}\right)^2 \delta^P(\hat{n}) > \left(\frac{c\alpha k}{r - c\tau}\right)^2$. The required vehicle capacity decreases.

When customers do not substitute between time windows, the demand density per time window remains constant, and each time window contributes the same profit margin. Proposition 3.4.1 characterizes the demand density for this margin to be positive. Naturally, this density decreases in unit revenues r and increases in the unit service time τ. Also note that it is independent of $\hat{n}$. Thus, for any demand density above the indicated threshold, increasing the number of time windows is always profitable. It is then the vehicle capacity, rather than the underlying economics that limits the total shift length. Conversely, if the demand density is too low to yield a positive contribution margin for a single time window, expanding the assortment does not benefit profitability. In this case, the focus should instead be on increasing the contribution margin – either by raising the demand density or by improving the net margin per order, $r - c\tau$.

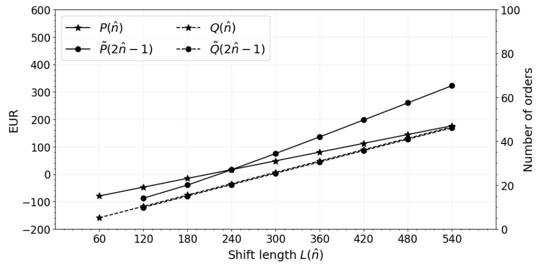
Interestingly, Proposition 3.4.2 reveals that overlapping time windows increase profits only if the additional demand can offset the additional time constraints introduced within the same time frame, which require more tours with fewer orders. Thus, the trade-off lies between higher demand and increased operational complexity due to stricter constraints, an effect also observed in Proposition 3.3.2.



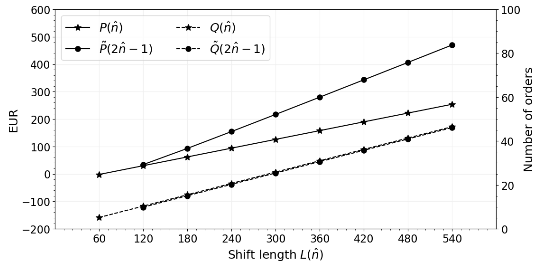
(a) Trade-off with $f = 50$



(b) Trade-off with $f = 10$



(c) Profit and required capacity with $f = 50$



(d) Profit and required capacity with $f = 10$

Figure 3.11.: Profitability of adjusting the number of time windows for varying number of consecutive time windows $\hat{n} \in \{1, \ldots, 9\}$ each of length $l = 60$ ($r = 9.0$, $c = 0.5$, $\tau = 8.0$, $\alpha = 2.0$, $k = 0.57$, $R = 100$, $\hat{N} = 10$)

Figure 3.11 illustrates numerical examples of the profit conditions stated in Proposition 3.4 (Figures 3.11a and 3.11b), as well as the trajectories of total profit and required vehicle capacity (Figures 3.11c and 3.11d) for a given demand density of $\frac{\hat{N}}{R} = 0.1$ per time window in a delivery region of size $R = 100$, and a varying number of consecutive time windows $\hat{n} \in \{1, \ldots, 9\}$ each of length $l = 60$ minutes. Figures 3.11c and 3.11d show that for consecutive time windows, the demand density is sufficiently large to ensure a positive contribution margin, leading to an increase in profit as additional time windows are added. In contrast, the addition of overlapping time windows becomes profitable only after reaching a certain shift length. Specifically, when comparing $P(\hat{n}+1)$ and $\widetilde{P}(2\hat{n}-1)$ in Figure 3.11c, we observe that only with a shift length of at least 360 minutes does adding overlapping options yield higher profit than extending the shift by one additional time window. However, extending the delivery shift requires more vehicle capacity, whereas the vehicle capacity requirement decreases slightly when overlaps are added within the same delivery shift.

Figures 3.11a and 3.11b depict that overlapping time windows require a larger contribution per time window, which decreases as the number of consecutive time windows increases. This explains why we see that the addition of overlapping time windows only becomes profitable after reaching a certain shift length. With fewer time windows, the attracted demand is insufficient to offset the additional fixed costs.

3.7. Conclusion

In grocery home delivery, designing an effective set of delivery time windows for customers to choose from is critical. The design task involves determining the number, length, and overlap of time windows, so as to balance customer preferences and operational efficiency. The strategic importance of this decision is evident in its impact on revenues, fulfillment costs, and vehicle capacity planning. However, it remains an under-researched area in the academic literature. Our work aims to address this gap by developing an analytical model that captures the trade-offs inherent in time window assortment design. By analyzing key factors influencing profitability and operational efficiency, the study seeks to provide managerial insights for aligning time window assortments with market conditions, enhancing the economic viability of online grocery delivery.

We provide accessible analytical expressions that identify which time window assortments can achieve profitability under varying market conditions. This includes minimum time window lengths, break-even demand thresholds, and conditions under which maximum vehicle sizes and vehicle tour utilization become constraining factors. Endogenizing demand, we show that for assortment-insensitive customers, service providers should offer the longest time window that still ensures efficient tour utilization, and there exists an optimal number of consecutive time windows that balances the key trade-off between clustering and spreading demand. For assortment-sensitive customers, there exists an optimal time window length that balances demand attraction and required tour frequency, and adding time windows (whether consecutive or overlapping) increases profit, provided the demand contribution is sufficiently large.

As we are the first to present a strategic model to support time window assortment design for attended home delivery, we see several directions for future research. First, it is interesting to study the interaction between strategic assortment decisions and tactical and operational offering and pricing decisions to smoothen demand over time and exploit customers' willingness to pay. Second, our analysis prioritizes profitability as a key performance measure, emphasizing economic viability. Future studies can extend the assortment evaluation model to include additional performance metrics, such as environmental sustainability (e.g., resource consumption and pollution) and social sustainability (e.g., working conditions and hours).

Chapter IV

Time Window Assortment Design with Stochastic Demand: The Value of Overlapping Time Windows

Abstract

Online grocery services require customers to be present at the time of delivery, making the design of time window assortments, including choices between many or few, long or short, overlapping or non-overlapping options, a relevant task. Despite growing interest in time window assortment design, the specific value of overlapping time windows remains not fully understood. Building on an established evaluation model that assesses the impact of time window assortments on both demand and delivery efficiency, we extend the framework to account for demand variability across time windows. This enhancement enables a detailed analysis of time window assortments under realistic customer choice behavior. In particular, we evaluate how overlapping time windows can help mitigate inefficiencies caused by demand peaks. We analytically derive necessary and sufficient conditions, based on realized demand, for overlapping time windows to reduce expected delivery costs compared to their consecutive counterparts. We then test these conditions and examine interaction effects of key parameters using Monte Carlo estimates. Our insights contribute significantly to a better understanding of how overlapping time windows affect delivery efficiency and under what operational conditions and demand patterns it is advisable for service providers to offer them.

4.1. Introduction

The rapid expansion of online grocery delivery has transformed how consumers shop for essentials, offering convenience while introducing logistical challenges for retailers. Unlike many e-commerce products, groceries are perishable, often require refrigeration, and must be delivered when customers are available to receive them. To coordinate deliveries efficiently, service providers commonly rely on structured time window assortments, allowing customers to select a preferred time window.

However, designing these time window assortments is far from straightforward. Providers must determine not only the number and length of available options but also if they should overlap. Overlapping time windows can boost demand and improve delivery capacity utilization, but they may also introduce inefficiencies in vehicle routing (Waßmuth et al., 2025). Moreover, demand uncertainty further complicates planning, as customer time window choices are subject to structural heterogeneity and statistical fluctuations. This paper examines the impact of overlapping time windows in the context of demand uncertainty.

Overlapping time windows are a common feature in demand management for attended home delivery, both in practice and research. Several studies have incorporated overlapping time windows to evaluate their demand management approaches (e.g., N. Agatz et al., 2011; Koch & Klein, 2020; Lang et al., 2021; Truden et al., 2022). Yet, the impact of overlapping time windows on demand and fulfillment has remained largely unexplored.

Waßmuth et al. (2025) were the first to analyze this relationship, showing that overlaps increase profit when they attract sufficient additional demand, but reduce profit when customers merely substitute between the offered options. That analysis, though, assumes demand to be balanced across time windows, which may not fully capture the strategic potential of overlapping assortments. In practice, demand across time windows is often uneven (Amorim et al., 2024), leading to inefficiencies such as lost demand or underutilized vehicle tours. While tactical and operational demand management can help mitigate these imbalances, they often incur additional costs, such as offering discounts (e.g., Campbell & Savelsbergh, 2006) or restricting time window availability (e.g., N. Agatz et al., 2011). Rather than addressing imbalances reactively for a given time window assortment, we propose a proactive approach: incorporating demand variability already during time window assortment design. Specifically, we argue that overlaps can serve as a hedge against intra-assortment demand variability.

This paper investigates the strategic value of offering overlapping time windows in attended home delivery, focusing on their cost performance under demand uncertainty. Building on the analysis of Waßmuth et al. (2025), we extend the scope to account for statistical fluctuations and structural differences in time window popularity. We compare overlapping and consecutive time window assortments in terms of their impact on delivery costs, isolating operational effects from revenue considerations. By examining how key problem parameters shape performance, we address the central research question: under what conditions do overlapping time windows offer advantages in delivery efficiency over their consecutive counterparts?

To achieve this, we extend the assortment evaluation model of Waßmuth et al. (2025) by representing demand across time windows as a joint random vector with a fixed total demand volume. We analytically characterize best- and worst-case cost performance and derive necessary and sufficient conditions under which overlapping time windows outperform their consecutive counterparts at the level of individual demand realizations. These insights are then translated into hypothesized performance criteria at the stochastic decision level. To validate and refine these criteria, we conduct a Monte Carlo simulation study that estimates expected performance across a comprehensive range of practically relevant parameter settings. Our contribution lies in assessing the operational value of overlapping time windows under assortment-dependent demand uncertainty and in offering actionable guidance on when overlaps improve delivery efficiency, thereby providing decision-makers with insights to inform strategic time window design.

The remainder of this paper is organized as follows. In Section 4.2, we review the related literature. We present the time window assortment evaluation model with stochastic demand in Section 4.3, which includes the statement of key assumptions and notation. In Section 4.4, we analytically assess cost performance on the level of individual demand realizations and in Section 4.5, we present our computational Monte Carlo study. We summarize our results and conclude with managerial implications in Section 4.6. All proofs are provided in Appendix B.

4.2. Related Literature

The management of delivery time windows is a central element of demand management in attended home delivery and has received growing attention in the literature (e.g., Cordeau et al., 2023; Fleckenstein et al., 2023; Waßmuth et al., 2023). Waßmuth et al. (2025) were the first to frame time window assortment design as a strategic planning problem, employing a continuous approximation model to derive analytical results under homogeneous conditions. However, that paper does not capture the potential of overlapping time windows to hedge against demand uncertainty. Earlier simulation-based studies explore relevant delivery configurations, but treat demand as exogenous (Boyer et al., 2009; Lin & Mahmassani, 2002; Punakivi & Saranen, 2001). We extend the model introduced by Waßmuth et al. (2025) by incorporating endogenous, stochastic demand and evaluating the value of overlapping time windows under assortment-dependent demand uncertainty.

Other literature on demand management primarily focuses on operational and tactical decisions, such as time window offering and pricing (e.g., Campbell & Savelsbergh, 2006; Campbell & Savelsbergh, 2005), as well as service differentiation and capacity control (e.g., Cleophas & Ehmke, 2014; Hernandez et al., 2017). A central challenge in this context is demand uncertainty: at the operational level, decisions must be made without knowing which customers will arrive next, creating opportunity costs (Fleckenstein et al., 2025), while at the tactical level, service providers must forecast aggregate demand with limited information. Most existing work either assumes deterministic forecasts (e.g., N. Agatz et al., 2011), formulates the problem as a scenario-based stochastic optimization model (e.g., Spliet & Gabor, 2015), or applies sampling-based and predictive solution approaches, as summarized in the recent methodological review by Fleckenstein et al. (2023). Our approach applies Monte Carlo sampling to estimate the expected performance of different time window assortments under stochastic demand.

The design of time window assortments is closely related to the field of assortment planning, which focuses on selecting a set of products or services to maximize revenue or service quality subject to operational constraints (Heger & Klein, 2024; Kök et al., 2015). A central element of this field is customer choice modeling, which captures how customers respond to the available options. Feng, Shanthikumar, and Xue (2022) provide a comprehensive review of commonly used choice models and their estimation methods. In attended home delivery, the offered time window assortment directly influences customer preferences, as customers may shift their demand from unavailable or less attractive options to other time windows within the assortment. Modeling these transitions is essential to capture substitution effects between different time window assortments. We simulate demand for consecutive assortments using a multinomial distribution and apply exogenous transition rates to overlapping options, enabling us to analyze how customer transitions influence the operational performance of overlapping versus consecutive time window assortments.

4.3. Assortment Evaluation Model with Stochastic Demand

We consider an e-grocery retailer that lets customers select a delivery time window from a menu of options, which we refer to as the time window assortment $T = (l, n, o)$. This assortment is characterized by the time window length l, the number of time windows n, and a categorical indicator o specifying whether the time windows are consecutive or overlapping.

In the following, we outline the key assumptions (Section 4.3.1) and present our demand (Section 4.3.2) and delivery model (Section 4.3.3) within the framework for time window assortment evaluation introduced in Waßmuth et al. (2025). Table 4.1 provides a summary of the main notation.

4.3.1. Assumptions

The retailer uses operational delivery shifts, and each time window assortment is associated with a particular shift. Moreover, they operate from a single delivery hub for a certain circular delivery region. Deliveries are affected by the time window constraints imposed by the time window assortment and the projected demand for that assortment. Our model builds on the following key assumptions.

Assumption 4.1. *Delivery shifts are independent.*

We treat delivery shifts as independent. This implies that there are no interactions between shifts neither in terms of demand nor delivery operations (Waßmuth et al., 2025).

Assumption 4.2. *Time window assortments, customer locations, orders, and vehicle tours are homogeneous.*

As in Waßmuth et al. (2025), this assumption implies identical gross margins across time windows and uniform operational conditions, which are characteristic of the continuous approximation method. However, while Waßmuth et al. (2025) additionally assume homogeneous demand, meaning that total demand is evenly distributed across the assortment's time windows, our model relaxes this assumption by allowing demand to vary between time windows. This variation captures both statistical fluctuations and structural differences in customer preferences. As a result, we are able to analyze the implications of overlapping time windows under demand variability.

Assumption 4.3. *The number of vehicle tours is set to cover the maximum demand across all time windows, thereby preventing any lost demand.*

The service provider determines the number of tours (i.e., vehicles used) after demand is realized, assuming a sufficiently large fleet to ensure that all demand is served. This assumption removes the need to model lost sales or unserved customers. However, since demand can vary across the time windows within a delivery shift, meeting peak demand may lead to idle driver time during periods of lower demand.

Assumption 4.4. *The total demand is independent of the time window assortment.*

This assumption implies that total demand is independent of both the number and length of time windows. For example, when overlapping time windows are introduced to a given consecutive assortment, the total number of orders remains unchanged. Customers reallocate their choices across the expanded set of options based on their preferences, altering only the distribution of demand across time windows while holding the overall demand volume constant. This modeling choice allows us to isolate the cost implications of the assortment design, especially those arising from the introduction of overlaps. The impact of time window length and number on demand volume and overall profit has already been demonstrated by Waßmuth et al. (2023) and is therefore not the focus of this study.

Taken together, these assumptions allow us to abstract away from revenue considerations and focus solely on delivery costs. They enable a clean analysis of how the introduction of overlapping windows affects delivery costs under variable demand distributions.

4.3.2. Demand Model

The assortment evaluation model presented in Waßmuth et al. (2025) includes a demand model that assumes demand to be homogeneous, i.e., orders are evenly distributed across the offered time windows. In this paper, we consider both statistical fluctuations in demand and structural differences in time window popularity.

We model the demand across all time windows in assortment T as a joint random vector $\boldsymbol{\mathcal{N}}^T = (\mathcal{N}_1^T, \mathcal{N}_2^T, \ldots, \mathcal{N}_n^T)$ that follows a probability distribution P_D constrained such that the total demand equals a fixed value D. A realization of the random vector is denoted by $\mathbf{N}^T = (N_1^T, N_2^T, \ldots, N_n^T)$, and by construction, the sum of the individual components equals the total demand: $\sum_{i=1}^n N_i^T = D$.

4.3.3. Delivery Model

Consider the delivery model introduced in Waßmuth et al. (2025). Removing the homogeneous demand assumption alters the model in two key ways. First, the workload must be determined separately for each time window to capture demand variation. Second, a decision rule is required to determine the number of tours to deploy from a sufficiently large delivery fleet once demand is realized. As specified in Assumption 4.3, we set the number of tours to match the workload of the peak time window, hereafter referred to as the *bottleneck workload*. As a result, vehicles may experience idle time during lower-demand time windows.

In the following, we define our performance metric, the *expected total delivery costs*, for consecutive time windows and overlapping time windows, respectively.

Consecutive Time Windows

We consider time window assortments $T \in \mathcal{T} = \{(l, n, o) \mid l \in \mathbb{R}^+, n \in \mathbb{N}^+, o = \text{con.}\}$, representing n consecutive time windows of length l, resulting in a total delivery shift of length $L(T) := nl$.

Following Assumption 4.3, we prevent any lost demand by ensuring that the number of vehicle tours is sufficient to serve all demand distributed according to $\boldsymbol{\mathcal{N}}^T$ within the given time windows. Consequently, the number of required tours is determined by the bottleneck workload, i.e., the maximum workload across the time windows:

$$v(T, \boldsymbol{\mathcal{N}}^T) := \frac{1}{l} \max_{i=1,\dots,n} w_i(\boldsymbol{\mathcal{N}}^T), \tag{4.1}$$

where $w_i(\boldsymbol{\mathcal{N}}^T) := \tau \mathcal{N}_i^T + \alpha k \sqrt{R \mathcal{N}_i^T}$ denotes the workload associated with time window i.

Both the individual workloads and the resulting number of tours are *stochastic*, as they depend on the random demand vector $\boldsymbol{\mathcal{N}}^T$, whose realization is only known after the order cut-off preceding the delivery shift. Thus, we define the *expected bottleneck workload* for assortment $T \in \mathcal{T}$ as

$$W(T) := \mathbb{E}_{\boldsymbol{\mathcal{N}}^T} \left[\max_{i=1,\dots,n} w_i(\boldsymbol{\mathcal{N}}^T) \right]. \tag{4.2}$$

The bottleneck workload is convex in the demand vector, as it is defined as the point-wise maximum of concave workload functions. With constant total demand, increasingly imbalanced demand distributions lead to higher bottleneck workloads. Figure 4.1 illustrates this effect. Interestingly, the concave shape of the individual workload functions appears almost linear. When one time window dominates the others, the associated demand becomes large enough that the linear service time term outweighs the square-root routing term, making the workload behave approximately linearly.

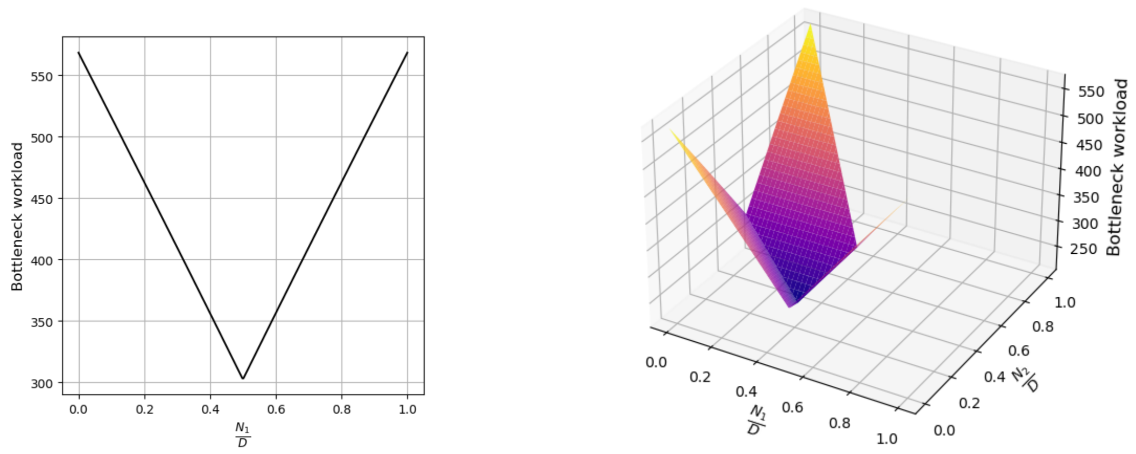
By the linearity of expectation, the *expected total delivery cost*, defined as the product of the cost per tour and the number of tours, is

$$C(T) := \mathbb{E}_{\boldsymbol{\mathcal{N}}^T} \left[C^t(T) \cdot v(T, \boldsymbol{\mathcal{N}}^T) \right] = \frac{C^t(T)}{l} \cdot W(T), \tag{4.3}$$

where $C^t(T) = cL(T) + F$ represents the cost per tour, with fixed costs $F = f + c\alpha \frac{4}{3}\sqrt{\frac{R}{\pi}}$.

Corollary 4.1. *Consider the expected total delivery cost $C(T)$ of an assortment $T \in \mathcal{T}$, as defined in Equation* (4.3)*, and let the demand vector $\boldsymbol{\mathcal{N}}^T$ follow a distribution P_D with fixed total demand volume D. Then, for any realization $\mathbf{N}^T$, a higher demand in a single time window increases the total delivery cost.*

This result forms the basis for all subsequent analyses. It follows directly from the time-based cost function defined in Equation (4.3) and the properties of the bottleneck



(a) $n = 2$ consecutive time windows **(b)** $n = 3$ consecutive time windows

Figure 4.1.: Illustration of the bottleneck workload for every potential demand distribution when the total demand is fixed at $D = 60$ ($\tau = 8$, $\alpha = 2$, $k = 0.57$, $R = 100$)

workload. Delivery costs depend on the number of vehicle tours covering the entire delivery shift, regardless of whether vehicles are driving or waiting. The bottleneck workload results from the no lost sales assumption (Assumption 4.3), which ensures that all demand must be served, regardless of how it is distributed across time windows. When demand is uneven, the number of required tours is determined by the peak load in the most heavily demanded window. As a result, assortments that lead to more balanced demand distributions reduce delivery costs and are therefore more cost-efficient.

Overlapping Time Windows

Consider assortments $T \in \widetilde{\mathcal{T}} = \{(l, n, o) \mid l \in \mathbb{R}^+, n = 2a + 1, a \in \mathbb{N}^+, o = \text{ovl.}\}$ that consist of n overlapping time windows of length l. We divide these time windows into $n + 1$ consecutive delivery intervals of length $\frac{l}{2}$, which results in a delivery shift length of $\widetilde{L}(T) := \frac{n+1}{2} l$.

Since the above route approximation applies to consecutive periods, we use the same demand allocation mechanism as in Waßmuth et al. (2025) to convert realized demand across overlapping time windows into demand over consecutive intervals. Figure 4.2 illustrates this allocation process using an example time window assortment.

Formally, for a given demand realization $\mathbf{N}^T = (N_1^T, N_2^T, \ldots, N_n^T)$, we model demand allocation using weights $\boldsymbol{\eta} \in [0, 1]^n$ representing the fraction of the time window demand

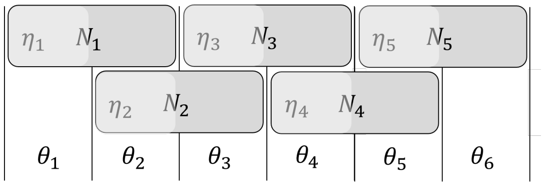


Figure 4.2.: Graphical representation of allocation decisions for an assortment of $n = 5$ overlapping time windows

that is allocated to the first half of the corresponding time window. We define allocation functions $\boldsymbol{\theta} = \big(\theta_i(\boldsymbol{\eta}, \mathbf{N}^T)\big)_{i=1,\dots,n+1}$ that determine the demand in the corresponding consecutive delivery intervals, where

$$\theta_i(\boldsymbol{\eta}, \mathbf{N}^T) := \begin{cases} \eta_i N_i^T, & i = 1, \\ (1-\eta_{i-1})N_{i-1}^T + \eta_i N_i^T, & i = 2, \dots, n, \\ (1-\eta_{i-1})N_{i-1}^T, & i = n+1. \end{cases} \tag{4.4}$$

The total demand volume remains unchanged at $\sum_{i=1}^{n} N_i^T = \sum_{i=1}^{n+1} \theta_i(\boldsymbol{\eta}, \mathbf{N}^T) = D$, and the realized workload in delivery interval i becomes $\widetilde{w}_i(\boldsymbol{\eta}, \mathbf{N}^T) = \tau\theta_i(\boldsymbol{\eta}, \mathbf{N}^T) + \alpha k\sqrt{R\theta_i(\boldsymbol{\eta}, \mathbf{N}^T)}$.

Given this allocation mechanism, we conclude that demand can be optimally allocated to minimize the bottleneck workload across the $n + 1$ delivery intervals of length $\frac{l}{2}$. This, in turn, reduces the required number of tours and lowers overall delivery costs, emphasizing the added flexibility that overlapping time windows provide in delivery planning.

As a result, estimating the operational performance of overlapping time windows corresponds to solving a *two-stage stochastic optimization problem* where the first stage *evaluates* operational performance for a fixed assortment $T \in \widetilde{\mathcal{T}}$, while the second stage *optimizes* the demand allocation based on observed demand realizations $\mathbf{N}^T$.

Formally, the *expected number of vehicle tours* is expressed as

$$\widetilde{v}(T) := \frac{2}{l}\mathbb{E}_{\boldsymbol{\mathcal{N}}^T}\left[R(T, \boldsymbol{\mathcal{N}}^T)\right], \tag{4.5}$$

Table 4.1.: Summary of notation used in Chapter IV

General Model Input	
α	Driving speed [min/km]
c	Variable cost per time unit [EUR/min]
D	Total demand volume [orders]
f	Vehicle-related fixed cost per tour [EUR]
F	Fixed cost per tour [EUR]
k	Road network factor
R	Surface of the delivery region [km^2]
τ	Delivery service time per order [min]
Assortment Input	
l	Time window length [min]
n	Number of time windows
o	Consecutive/overlapping flag
$T \in \mathcal{T} \cup \widetilde{\mathcal{T}}$	Time window assortment
Demand Uncertainty	
$\boldsymbol{\mathcal{N}}^T = (\mathcal{N}_1^T, \mathcal{N}_2^T, \ldots, \mathcal{N}_n^T)$	Random demand vector [orders]
$\mathbf{N}^T = (N_1^T, N_2^T, \ldots, N_n^T)$	Demand realization [orders]
Performance Output	
C^t	Delivery cost per tour [EUR]
$\boldsymbol{\eta}$	Allocation weights
L	Delivery shift length [min]
v	Number of vehicle tours
w_i	Workload in time window $i = 1, \ldots, n$ [min]
W	Bottleneck workload [min]

where the expectation of the recourse function $R(T, \cdot)$ is taken with respect to the random demand vector $\boldsymbol{\mathcal{N}}^T$. For any realization $\mathbf{N}^T$, the recourse function

$$R(T, \mathbf{N}^T) := \min_{\boldsymbol{\eta} \in [0,1]^n} \left(\max_{i=1,\ldots,n+1} \widetilde{w}_i(\boldsymbol{\eta}, \mathbf{N}^T) \right) \tag{4.6}$$

represents the optimal bottleneck workload, where $\widetilde{w}_i(\boldsymbol{\eta}, \mathbf{N}^T)$ denotes the workload allocated to delivery interval i, given allocation $\boldsymbol{\eta}$ and demand realization $\mathbf{N}^T$.

To compare consecutive and overlapping assortments effectively, we define the relevant *expected bottleneck workload* for an overlapping assortment $T \in \widetilde{\mathcal{T}}$ as

$$\widetilde{W}(T) := 2 \cdot \mathbb{E}_{\boldsymbol{\mathcal{N}}^T}\left[R(T, \boldsymbol{\mathcal{N}}^T)\right]. \tag{4.7}$$

Consistent with the linear cost structure in Equation (4.3), the *expected total delivery cost* is then given by

$$\widetilde{C}(T) := \frac{\widetilde{C}^t(T)}{l} \cdot \widetilde{W}(T), \tag{4.8}$$

with a cost per tour of $\widetilde{C}^t(T) = c\widetilde{L}(T) + F$.

In the following, we establish an understanding of the structure of optimal second-stage solutions. Since we consider all parameter values to be positive, the optimal solution of Equation (4.6) is equivalent to the optimal solution of a *linear optimization problem* that minimizes the maximum demand, or *bottleneck demand*, across the consecutive delivery intervals, represented by

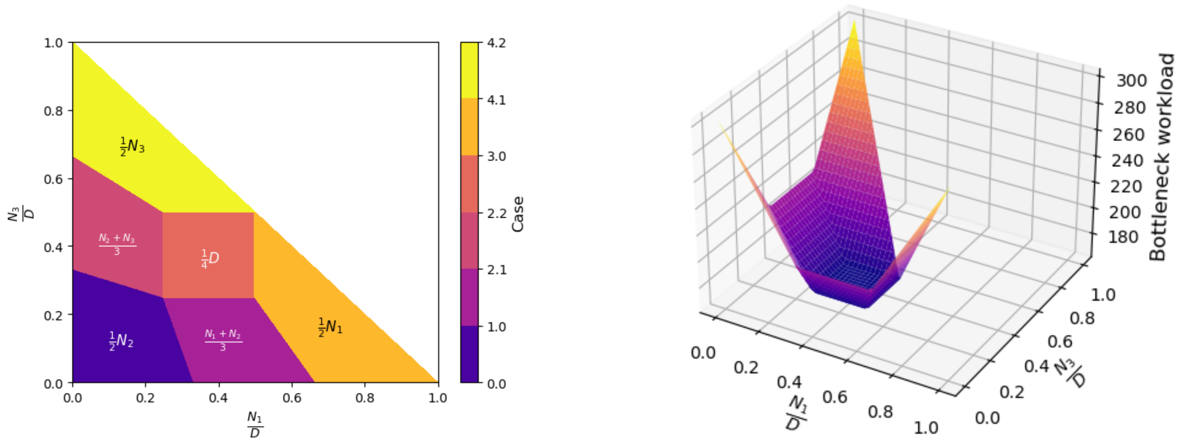
$$\begin{aligned} \min_{(Z, \boldsymbol{\eta})} \quad & Z \\ \text{s.t.} \quad & Z \geq \theta_i(\boldsymbol{\eta}, \mathbf{N}^T) \quad && i = 1, \ldots, n+1, \\ & \eta_i \in [0, 1] && i = 1, \ldots, n \end{aligned} \tag{4.9}$$

For general n-dimensional demand, Equation (4.9) can be efficiently solved numerically, but obtaining a closed-form solution is analytically intractable. To gain structural insights into the problem, however, we derive a closed-form solution for the special case of $n = 3$ overlapping time windows.

Lemma 4.1. *Consider a time window assortment $T \in \widetilde{\mathcal{T}}$ consisting of $n = 3$ overlapping time windows. Let $\mathbf{N} = (N_i \geq 0)_{i=1,\ldots,3}$ be a non-negative demand realization with total demand volume $D = \sum_{i=1}^{3} N_i > 0$. The optimal allocation of demand $\boldsymbol{\eta}^*$ yields the following expressions for the optimal bottleneck demand Z^*:*

Case	N_1	N_3	Z^*
1	$0 \leq \frac{N_1}{D} \leq \frac{1}{3}(1 - \frac{N_3}{D})$	$0 \leq \frac{N_3}{D} \leq \frac{1}{3}(1 - \frac{N_1}{D})$	$\frac{1}{2}N_2$
2.1	$\frac{1}{3}(1 - \frac{N_3}{D}) \leq \frac{N_1}{D} \leq \frac{2}{3}(1 - \frac{N_3}{D})$	$0 \leq \frac{N_3}{D} \leq \frac{1}{4}$	$\frac{1}{3}(N_2 + N_1)$
2.2	$0 \leq \frac{N_1}{D} \leq \frac{1}{4}$	$\frac{1}{3}(1 - \frac{N_1}{D}) \leq \frac{N_3}{D} \leq \frac{2}{3}(1 - \frac{N_1}{D})$	$\frac{1}{3}(N_2 + N_3)$
3	$\frac{1}{4} \leq \frac{N_1}{D} \leq \frac{1}{2}$	$\frac{1}{4} \leq \frac{N_3}{D} \leq \frac{1}{2}$	$\frac{1}{4}D$
4.1	$\max\left(\frac{2}{3}(1 - \frac{N_3}{D}), \frac{1}{2}\right) \leq \frac{N_1}{D} \leq 1 - \frac{N_3}{D}$	$0 \leq \frac{N_3}{D} \leq \frac{1}{2}$	$\frac{1}{2}N_1$
4.2	$0 \leq \frac{N_1}{D} \leq \frac{1}{2}$	$\max\left(\frac{2}{3}(1 - \frac{N_1}{D}), \frac{1}{2}\right) \leq \frac{N_3}{D} \leq 1 - \frac{N_1}{D}$	$\frac{1}{2}N_3$

Lemma 4.1 partitions the parameter space of the a demand vector $\mathbf{N} = (N_i \geq 0)_{i=1,\dots,3}$ into distinct regions within the 2-simplex, each reflecting a specific degree of balance in demand allocation. For example, *Case 3* represents the range of time window demands that corresponds to a fully balanced demand allocation. The remaining cases each result in the best possible balance, i.e., the lowest possible bottleneck demand.



(a) Valid parameter range and maximum demand per time window for optimal allocation decisions

(b) Bottleneck workload resulting from optimal allocation decisions ($D = 60$, $\tau = 8$, $\alpha = 2$, $k = 0.57$, $R = 100$)

Figure 4.3.: Illustration of the optimal allocation decisions introduced in Lemma 4.1

Figure 4.3 visualizes this result. Figure 4.3a presents the strategy plot, which partitions the parameter space of possible demand realizations $\mathbf{N}$ into regions corresponding to the optimal allocation strategies. It also visualizes the corresponding bottleneck demand Z^* across the consecutive delivery intervals. Figure 4.3b presents a numerical example of the bottleneck workload resulting from the optimal allocation strategy applied to all possible demand realizations $\mathbf{N}^T$ for a given total demand volume and a given set of delivery region characteristics. In accordance with the property stated in Corollary 4.1, it generally holds that the more evenly time window demand can be redistributed, the lower the relevant bottleneck workload, thereby reducing the number of tours and total delivery costs.

In conclusion, the analytical solution presented in Lemma 4.1, together with Corollary 4.1, provides strong evidence that the allocation flexibility enabled by overlapping time windows can lead to substantial efficiency gains compared to consecutive assortments. These findings offer valuable intuition that extends beyond the simplified setting and remains relevant for more complex problems with n-dimensional demand.

4.4. Theoretical Performance Assessment

The following analyses focus on comparing overlapping time window assortments with their corresponding consecutive assortment, based on the expected total delivery costs defined in Equations (4.3) and (4.8). Theoretically assessing the decision to introduce overlapping time windows is challenging because demand realizations are unknown at the time of assortment selection, and the inherent uncertainty is difficult to analyze analytically. Therefore, this section conducts an ex-post analytical assessment of the performance of consecutive and overlapping time window assortments for given demand realizations.

Understanding the conditions under which overlapping assortments lead to delivery cost reductions relative to consecutive assortments offers valuable insights for decision-making under uncertainty. To this end, we derive performance bounds based on best- and worst-case demand realizations and identify necessary and sufficient conditions under which overlapping assortments outperform their consecutive counterparts.

4.4.1. Best- and Worst-case Outcomes

To gain insight into the best- and worst-case outcomes under uncertain demand $\boldsymbol{\mathcal{N}}^T$, we derive upper and lower bounds on the stochastic bottleneck workloads that underlie the expectations defined in Equations (4.2) and (4.7). We focus on bottleneck workloads, as they capture the impact of demand uncertainty on total delivery costs.

For consecutive assortments $T \in \mathcal{T}$, the best-case demand realization occurs when demand is evenly distributed across all time windows, yielding a bottleneck demand of $\frac{D}{n}$. In contrast, the worst-case realization arises when the entire demand is concentrated in a single time window, resulting in a bottleneck demand of D. Consequently, the worst-case realization leads to a bottleneck demand that is n times higher than in the balanced case. This observation forms the basis of the following lemma.

Lemma 4.2. *Consider Equations* (4.2) *and* (4.7)*, let the corresponding demand vectors* $\boldsymbol{\mathcal{N}}^T$ *follow a distribution with fixed total demand volume* D*, and define* $w^D := \tau D + \alpha k\sqrt{RD}$*. We establish bounds on the relevant bottleneck workloads for consecutive and overlapping time windows, respectively.*

a) For an assortment $T \in \mathcal{T}$ *consisting of* n *consecutive time windows, the bottleneck workload is bound by*

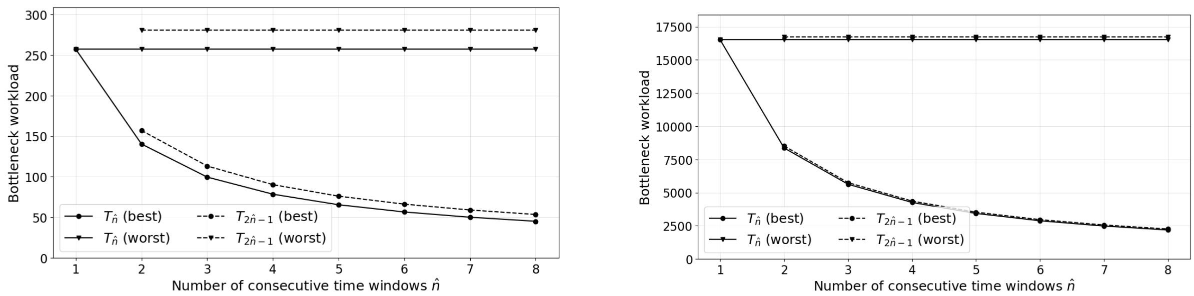
$$\frac{1}{\sqrt{n}}w^D - \left(\frac{1}{\sqrt{n}} - \frac{1}{n}\right)\tau D \leq \max_{i=1,\dots,n} w_i(\boldsymbol{\mathcal{N}}^T) \leq w^D$$

b) For an assortment $T \in \widetilde{\mathcal{T}}$ *consisting of* n *overlapping time windows, the bottleneck workload is bound by*

$$\frac{2}{\sqrt{n+1}}w^D - \left(\frac{2}{\sqrt{n+1}} - \frac{2}{n+1}\right)\tau D \leq 2 \cdot R(T, \boldsymbol{\mathcal{N}}^T) \leq \sqrt{2}w^D - \left(\sqrt{2} - 1\right)\tau D$$

Naturally, for consecutive time windows, best-case and worst-case scenarios yield identical performance when only a single time window is offered ($n = 1$). As the number of time windows increases ($n \to \infty$), the performance gap between the best-case and worst-case scenarios widens, driven by lower bottleneck workloads in the best-case, albeit at a decreasing rate. This is intuitive: a broader assortment increases demand variability across options (see also Zhang et al., 2022), which, under favorable conditions, can be exploited to improve cost efficiency. In the case of overlapping time windows, even small assortments ($n = 3$) exhibit a performance gap between the best-case and worst-case scenarios. This is because near-balanced demand can be effectively smoothed, whereas inefficiencies caused by concentrated demand can only be partially mitigated.

In Figure 4.4, we compare examples of the best- and worst-case performance of consecutive assortments (solid line) and corresponding overlapping assortments (dashed line)



(a) $D = 25$ **(b)** $D = 2000$

Figure 4.4.: Illustration of performance bounds for consecutive and overlapping time windows ($\tau = 8$, $\alpha = 2$, $k = 0.57$, $R = 100$)

depending on the assortment size. The consecutive assortment is denoted by $T_{\hat{n}} \in \mathcal{T}$, while the overlapping assortment is represented by $T_{2\hat{n}-1} \in \widetilde{\mathcal{T}}$, with each consisting of $\hat{n} \in \{1, \ldots, 8\}$ consecutive time windows and capturing a total demand volume of D. Under the best-case demand realization, the figures show that consecutive time windows yield a lower bottleneck workload, and thus lower costs, than overlapping ones, consistent with the findings of Waßmuth et al. (2025).

Interestingly, even under the worst-case scenario, consecutive time windows consistently outperform their overlapping counterparts. This indicates that the potential for demand redistribution through overlapping assortments is inherently limited: in certain instances, the added flexibility does not sufficiently compensate for the increased complexity caused by denser constraints within a fixed delivery shift, as reflected in our model.

Moreover, the best-case performance of overlapping time windows converges to that of consecutive ones as the assortment size n increases. In addition, when comparing a low total demand density of 0.25 customers per km^2 (Figure 4.4a) with a high density of 20 customers per km^2 (Figure 4.4b), we observe that both the best- and worst-case performances of overlapping assortments approach those of consecutive time windows as total demand increases. The following section examines these observations in greater detail.

4.4.2. When Overlapping Time Windows Are Cost-Efficient

We have seen that overlapping time windows underperform their consecutive counterparts in terms of best-case and worst-case delivery costs. Nonetheless, we argue that the flexibility offered by overlapping windows can reduce costs in scenarios with sufficient demand imbalances. To identify such cases, we derive necessary and sufficient conditions under which overlapping assortments outperform consecutive ones.

Consider a consecutive time window assortment $T_{\hat{n}} = (l, \hat{n}, \text{con.}) \in \mathcal{T}$ with demand realization $\mathbf{N}^{T_{\hat{n}}}$, and a corresponding overlapping assortment $T_{2\hat{n}-1} = (l, 2\hat{n}-1, \text{ovl.}) \in \widetilde{\mathcal{T}}$ with demand realization $\mathbf{N}^{T_{2\hat{n}-1}}$. Both assortments contain $\hat{n}$ consecutive time windows of length l, such that they yield the same delivery shift length: $L(T_{\hat{n}}) = \widetilde{L}(T_{2\hat{n}-1})$. As a result, the delivery cost per tour is identical for both assortments: $C^t(T_{\hat{n}}) = \widetilde{C}^t(T_{2\hat{n}-1})$.

Therefore, the overlapping assortment $T_{2\hat{n}-1}$ yields a lower total delivery cost than the consecutive assortment $T_{\hat{n}}$, i.e., $\widetilde{C}(T_{2\hat{n}-1}, \mathbf{N}^{T_{2\hat{n}-1}}) < C(T_{\hat{n}}, \mathbf{N}^{T_{\hat{n}}})$, if and only if it results in a lower bottleneck workload: $\widetilde{W}(T_{2\hat{n}-1}, \mathbf{N}^{T_{2\hat{n}-1}}) < W(T_{\hat{n}}, \mathbf{N}^{T_{\hat{n}}})$. Accordingly, the following theorem characterizes the extent of demand redistribution required for overlapping time windows to be cost-efficient.

Theorem 4.1. *Consider Equations* (4.3) *and* (4.8) *for any assortments* $T_{\hat{n}} \in \mathcal{T}$ *and* $T_{2\hat{n}-1} \in \widetilde{\mathcal{T}}$, *each comprising* $\hat{n}$ *consecutive time windows of length* l. *Let* $\mathbf{N}^{T_{\hat{n}}}$ *and* $\mathbf{N}^{T_{2\hat{n}-1}}$ *denote any respective demand realizations. Define the bottleneck demand* $\bar{N} := \max_{i=1,\ldots,\hat{n}} N_i^{T_{\hat{n}}}$ *and the allocated bottleneck demand* $\bar{\theta} := \max_{i=1,\ldots,2\hat{n}} \theta_i(\boldsymbol{\eta}^*, \mathbf{N}^{T_{2\hat{n}-1}})$ *under the optimal allocation* $\boldsymbol{\eta}^*$. *The overlapping assortment outperforms the consecutive assortment, i.e.,* $\widetilde{C}(T_{2\hat{n}-1}, \mathbf{N}^{T_{2\hat{n}-1}}) < C(T_{\hat{n}}, \mathbf{N}^{T_{\hat{n}}})$, *if and only if*

$$\frac{\bar{\theta}}{\bar{N}} < \frac{1}{2} - \frac{d(\bar{N})}{2\tau^2}\left(\sqrt{w^c(\bar{N})^2 + \tau^2} - w^c(\bar{N})\right) < \frac{1}{2},$$

with $d(\bar{N}) = \alpha k \sqrt{\frac{R}{\bar{N}}}$ *and* $w^c(\bar{N}) = \tau + d(\bar{N})$.

Theorem 4.1 establishes a threshold for the *normalized allocation* $\frac{\bar{\theta}}{\bar{N}}$, which determines whether the introduction of overlapping time windows leads to a reduction in total delivery costs. It also explains the weaker worst-case performance of such assortments, stated in Lemma 4.2: in this case, the allocated bottleneck demand $\bar{\theta}$ equals exactly half of the bottleneck demand $\bar{N}$ realized under consecutive time windows, whereas it would need to be strictly lower, by an amount specified in the lemma, to improve cost

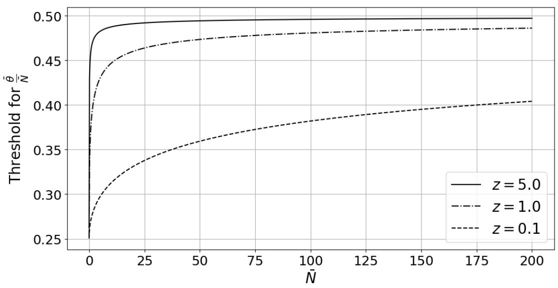


Figure 4.5.: Illustration of the required redistribution level for $\bar{N} \in (0, \dots, 200)$ with $\tau = z\alpha k\sqrt{R}$ ($\alpha = 2$, $k = 0.57$, $R = 100$)

efficiency. Notably, the derived threshold enables the evaluation of overlapping time windows for any potential demand realization resulting from assortment expansion, and it depends only on four delivery region characteristics: the service time τ, the region size R, the driving speed α, the road network factor k.

Figure 4.5 visualizes this threshold for varying values of $\bar{N}$ and different proportions between service time τ and routing effort $\alpha k\sqrt{R}$, denoted by $z := \frac{\tau}{\alpha k\sqrt{R}}$. The figure illustrates a key insight: the smaller the bottleneck demand $\bar{N}$ under consecutive time windows, the lower the normalized allocation $\frac{\bar{\theta}}{\bar{N}}$ must be for the overlapping assortment to yield lower costs. When service time dominates routing time (e.g., $z > 1$), the bound quickly approaches $\frac{1}{2}$, and stricter conditions on $\frac{\bar{\theta}}{\bar{N}}$ arise only for small values of $\bar{N}$. In contrast, when routing time becomes more significant, lower normalized allocations are required even for larger values of $\bar{N}$.

In conclusion, Theorem 4.1 establishes a tractable necessary and sufficient condition under which overlapping time windows reduce total delivery costs, given a pair of corresponding demand realizations. Moreover, even at the time of the assortment decision, when exact demand information is unavailable, the condition provides valuable insight into when cost savings from overlapping time windows are more attainable. That is, when the derived threshold assumes a higher value, thereby relaxing the constraint on the normalized allocation $\frac{\bar{\theta}}{\bar{N}}$. Accordingly, the following conjecture summarizes three scenarios that are expected to promote cost savings through overlapping time windows.

Conjecture 4.1. *The following three scenarios are hypothesized to support the condition in Theorem 4.1, under which overlapping assortments yield cost savings compared to their consecutive counterpart: (i) a pronounced imbalance in demand across the consecutive* *assortment, (ii) a high overall demand volume, and (iii) service time dominating routing time.*

These scenarios all relate to the right-hand side of the condition in Theorem 4.1. We now focus on the left-hand side, the normalized demand $\frac{\bar{\theta}}{\bar{N}}$. Specifically, we quantify the maximum demand reallocation under the allocation mechanism defined in Equation (4.4).

Lemma 4.3. *Let $T \in \widetilde{\mathcal{T}}$ be any overlapping assortment, $\mathbf{N}^T$ any corresponding demand realization, and consider the allocation mechanism defined in Equation* (4.4)*. Let $\hat{N} := \max_{i=1,\dots,n} N_i^T$ denote the bottleneck demand, and let $\bar{\theta} := \max_{i=1,\dots,n+1} \theta_i(\boldsymbol{\eta}^*, \mathbf{N}^T)$ be the allocated bottleneck demand under optimal allocation $\boldsymbol{\eta}^*$. Then, the allocation procedure cannot reduce the bottleneck demand by more than half:*

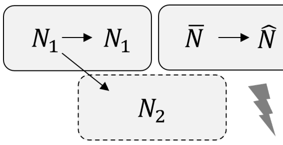
$$\frac{\bar{\theta}}{\hat{N}} \geq \frac{1}{2}.$$

Corollary 4.2. *Overlapping time windows can lead to a reduction in delivery costs only if customer choice across time windows results in a decrease of the bottleneck demand relative to the consecutive case, i.e., if $\hat{N} < \bar{N}$.*

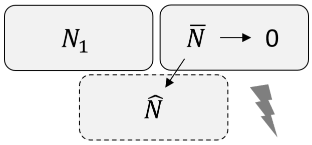
The corollary follows directly from Theorem 4.1 and Lemma 4.3, establishing a necessary condition for the cost-efficiency of overlapping time windows that relates to customer behavior. In other words, even when the favorable scenarios stated in Conjecture 4.1 are present, cost savings from overlapping time windows occur only if customers self-select across time windows in a way that reduces the bottleneck demand relative to the consecutive case.

Figures 4.6a to 4.6c illustrate this condition across three extreme cases. In both Figure 4.6a and Figure 4.6b, the bottleneck demand remains at the same level, $\hat{N} = \bar{N}$, implying that the condition in Theorem 4.1 cannot be satisfied. Among the illustrated cases, only the demand pattern in Figure 4.6c allows for potential cost reduction through the introduction of an overlapping time window, although it does not guarantee it.

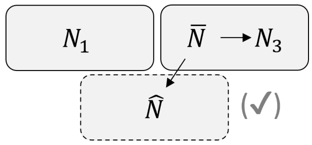
Building on this, Figures 4.6d and 4.6e illustrate the complete set of customer substitutions between the consecutive and overlapping assortments, given an initial demand realization that matches the special case outlined in Figure 4.6c. We indicate scenarios for which adding overlapping time windows does or does not reduce costs, using



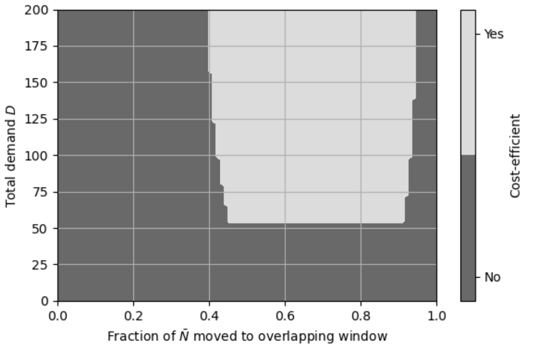
(a) Some low demand is drawn to the overlapping time window



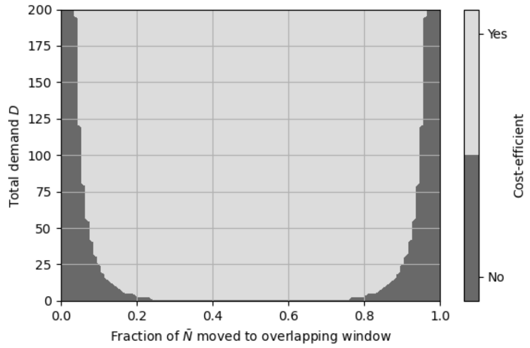
(b) All bottleneck demand is drawn to the overlapping time window



(c) Some bottleneck demand is drawn to the overlapping time window



(d) Cost-efficiency map given an initial imbalance with a bottleneck demand of $\bar{N} = 0.55D$



(e) Cost-efficiency map given an initial imbalance with a bottleneck demand of $\bar{N} = 0.95D$

Figure 4.6.: Examples of cost-efficiency results when transitioning from $\hat{n} = 2$ consecutive to $2\hat{n} - 1 = 3$ overlapping time windows for varying customer behavior and a fixed service-to-routing time ratio of $z = 0.7$ ($\alpha = 2$, $k = 0.57$, $R = 100$)

two example demand realizations across varying levels of total demand D and a fixed service-to-routing time ratio $z = 0.7$: one relatively balanced, with a bottleneck demand of $\bar{N} = 0.55D$ (Figure 4.6d), and the other highly imbalanced, with a bottleneck demand of $\bar{N} = 0.95D$ (Figure 4.6e).

As expected, the relatively balanced case leaves less room for the overlapping time windows to be cost-efficient, only for higher total demand and substitution patterns that favorably support the allocation mechanism in sufficiently relieving bottleneck demand. In contrast, the overlapping assortment outperforms the consecutive counterpart in nearly all substitution scenarios in the imbalanced case, particularly at higher total demand levels.

Concluding this section, we have established necessary and sufficient conditions at the demand realization level that determine when adding overlapping time windows reduces delivery costs compared to the corresponding consecutive assortment. At the assortment decision level, these results suggest that overlaps can create value when they are expected to facilitate effective demand redistribution. However, since our analysis focuses solely on cost effects, we do not account for the potential of overlapping time windows to attract additional demand – a dimension addressed in Waßmuth et al. (2025).

4.5. Monte Carlo Simulation

Given the conditions for cost-efficient overlapping assortments outlined above, we now assess how likely these conditions are to hold under stochastic demand. To this end, we employ Monte Carlo simulation to estimate and compare the expected total delivery costs of overlapping time window assortments and their corresponding consecutive alternatives. The comparison spans various combinations of key parameters related to customer behavior, total demand volume, and the ratio of service time to routing time. In Section 4.5.1, we outline the experimental design of our Monte Carlo simulation. Section 4.5.2 presents and discusses the corresponding numerical results.

4.5.1. Experimental Design

Consider a time window assortment $T_{\hat{n}} = (l, \hat{n}, \text{con.}) \in \mathcal{T}$, consisting of $\hat{n}$ consecutive time windows, each of length l. Let $T_{2\hat{n}-1} = (l, 2\hat{n}-1, \text{ovl.}) \in \widetilde{\mathcal{T}}$ be a corresponding overlapping assortment that covers the same time span using $2\hat{n}-1$ overlapping time windows of length l. These assortments yield the same total demand D, the same delivery shift length, $L(T_{\hat{n}}) = \widetilde{L}(T_{2\hat{n}-1})$, and the same delivery cost per tour, $C^t(T_{\hat{n}}) = \widetilde{C}^t(T_{2\hat{n}-1})$. Consequently, comparing the expected total delivery costs, $C(T_{\hat{n}})$ and $\widetilde{C}(T_{2\hat{n}-1})$, reduces to comparing the expected bottleneck workloads, $W(T_{\hat{n}})$ and $\widetilde{W}(T_{2\hat{n}-1})$. Ultimately, our objective is to estimate the expected relative difference between these bottleneck workloads.

To do so, we draw m samples from a consecutive demand distribution $\boldsymbol{\mathcal{N}}^{T_{\hat{n}}} \sim P_D$ that satisfies a fixed-sum property, with each realization denoted by $\mathbf{N}^{T_{\hat{n}},j}$ for $j = 1, \ldots, m$. We first introduce the specific probability distribution used in our analysis. Further, we introduce the specific customer behavior based on which each demand realization for the consecutive assortment is transformed into a corresponding realization for the overlapping assortment, denoted by $\mathbf{N}^{T_{2\hat{n}-1},j}$. For each sample and scenario that we introduce afterwards, we compute the corresponding bottleneck workloads, denoted by $W(T_{\hat{n}}, \mathbf{N}^{T_{\hat{n}},j})$ and $\widetilde{W}(T_{2\hat{n}-1}, \mathbf{N}^{T_{2\hat{n}-1},j})$, respectively.

We obtain the Monte Carlo estimator for the expected relative difference of bottleneck workloads as

$$\widehat{\Delta}_{W(\hat{n})} = \frac{1}{m}\sum_{j=1}^{m} 100 \times \left(\frac{\widetilde{W}(T_{2\hat{n}-1}, \mathbf{N}^{T_{2\hat{n}-1},j}) - W(T_{\hat{n}}, \mathbf{N}^{T_{\hat{n}},j})}{W(T_{\hat{n}}, \mathbf{N}^{T_{\hat{n}},j})}\right). \tag{4.10}$$

By the law of large numbers, the estimator $\widehat{\Delta}_{W(\cdot)}$ converges to the true expectation as $m \to \infty$. Note that evaluating each instance of $\widetilde{W}(T_{2\hat{n}-1}, \mathbf{N}^{T_{2\hat{n}-1},j})$ requires solving the linear optimization problem defined in Equation (4.9).

Probability Distribution

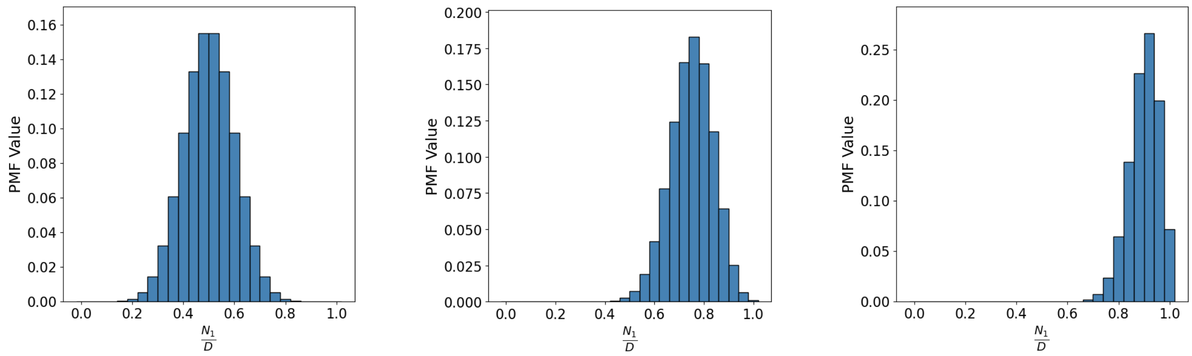
Consider any assortment $T = (l, n, o)$ consisting of n time windows. We model the corresponding demand vector $\boldsymbol{\mathcal{N}}^T$ using a multinomial distribution with total demand D and probability vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$, where $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$:

$$(\mathcal{N}_1^T, \mathcal{N}_2^T, \ldots, \mathcal{N}_n^T) \sim \text{Multinomial}(D, \mathbf{p}). \tag{4.11}$$

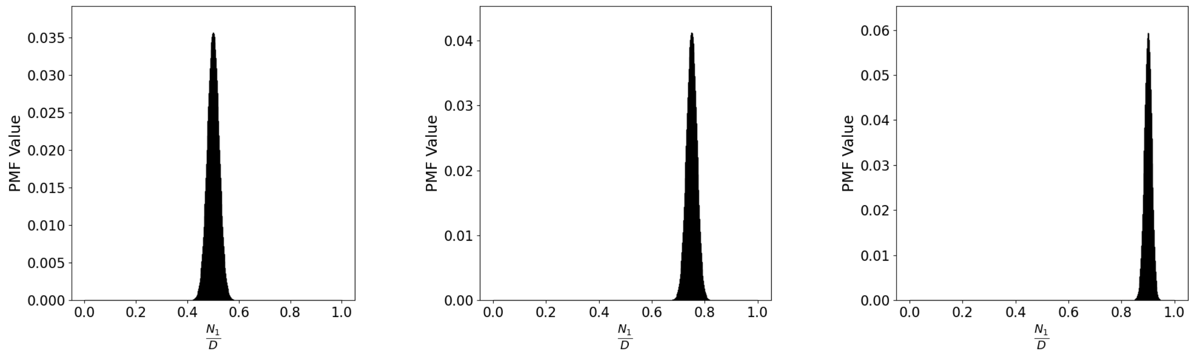
The multinomial distribution forms the basis of the Multinomial Logit (MNL) model, a widely used model for representing customer choice over discrete assortments, particularly in operations and revenue management contexts (Feng, Shanthikumar, & Xue, 2022). Due to its tractability and closed-form expressions for choice probabilities, the MNL model is frequently employed in assortment and price optimization (see e.g., K. Talluri & Van Ryzin, 2004; R. Wang, 2012). Despite its simplifying assumptions, it provides a practical foundation for modeling customer preferences in large-scale decision-making systems.

In our setting, the multinomial distribution captures the allocation of total demand D across the time windows in assortment T, under the assumption that each unit of demand independently selects time window i with probability p_i. Importantly, we do not consider an outside option, as we abstract from any revenue effects. That is, all customers are assumed to make a selection from within the offered time windows.

Formally, the probability mass function of the multinomial distribution is given by $P(\mathcal{N}_1^T = x_1, \ldots, \mathcal{N}_n^T = x_n) = \frac{D!}{x_1! \, x_2! \cdots x_n!} \prod_{i=1}^{n} p_i^{x_i}$, where $x_i \in \mathbb{Z}_{\geq 0}$ and $\sum_{i=1}^{n} x_i = D$. Figure 4.7 depicts this function for different demand volumes D and several probability vectors $\mathbf{p}$, using the example of $n = 2$ time windows. As the demand volume increases,



(a) $D = 25$ and $\mathbf{p} = (0.5, 0.5)$
(b) $D = 25$ and $\mathbf{p} = (0.75, 0.25)$
(c) $D = 25$ and $\mathbf{p} = (0.9, 0.1)$



(d) $D = 500$ and $\mathbf{p} = (0.5, 0.5)$
(e) $D = 500$ and $\mathbf{p} = (0.75, 0.25)$
(f) $D = 500$ and $\mathbf{p} = (0.9, 0.1)$

Figure 4.7.: Probability mass function of the Multinomial distribution for $n = 2$ time windows

the distribution becomes increasingly concentrated around its mean. In contrast, for small demand volumes, the observed allocations exhibit substantial variability due to the discrete nature of the multinomial distribution and the relatively high variance when the number of trials is low.

In our numerical study, we sample demand realizations $\mathbf{N}^{T_{\hat{n}},j}$ of consecutive assortments $T_{\hat{n}} = (l, \hat{n}, \text{con.}) \in \mathcal{T}$ from the multinomial distribution Multinomial$(D, \mathbf{p})$ with total demand D and probability vector $\mathbf{p} = (p_1, \dots, p_{\hat{n}})$.

Customer Behavior

Comparing consecutive and overlapping assortments $T_{\hat{n}}$ and $T_{2\hat{n}-1}$ requires accounting for customer substitution behavior when overlapping time windows are introduced. In

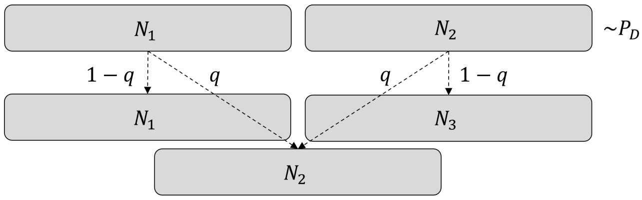


Figure 4.8.: Illustration of customer transitions for $\hat{n} = 2$ consecutive time windows

other words, we must capture how customer demand redistributes as the assortment expands from consecutive to overlapping windows. To ensure this, a sample for the overlapping assortment $\mathbf{N}^{T_{2\hat{n}-1},j}$ must be drawn conditional on a sample for the consecutive assortment $\mathbf{N}^{T_{\hat{n}},j}$, reflecting consistent customer choice behavior across both assortments.

Accordingly, we sample demand for the overlapping assortment using a two-step procedure. First, we draw a realization $\mathbf{N}^{T_{\hat{n}},j}$ of demand for the consecutive assortment, as described above. In the second step, we derive the corresponding realization for the overlapping assortment by applying a transformation function, $\mathbf{N}^{T_{2\hat{n}-1},j} := \text{subst}(\mathbf{N}^{T_{\hat{n}},j}, q)$, based on a transition parameter $q \in [0, 1]$. This parameter specifies the proportion of customers in each consecutive time window who are expected to shift to one of the adjacent overlapping time windows. Figure 4.8 illustrates the underlying logic of these customer transitions.

Scenario Definition

We carefully define the set of scenarios to test our hypotheses in Conjecture 4.1, to deepen our understanding of the impact of different demand distributions and transition rates, and to assess the impact of the time window assortment size. Throughout all subsequent analyses, we fix the delivery region size to $R = 100$ km^2, the driving speed to 30 km/h, ($\alpha = 2$ min/km), and the road network factor to $k = 0.57$.

First, we consider consecutive assortments $T_2 \in \mathcal{T}$ consisting of $\hat{n} = 2$ time windows, and corresponding overlapping assortments $T_3 \in \widetilde{\mathcal{T}}$ comprising $2\hat{n} - 1 = 3$ time windows. We evaluate three levels of the service-to-routing time ratio, $z \in \{0.1,\ 0.7,\ 1.5\}$, corresponding to service times τ of approximately 1, 8, and 17 minutes. We further consider three levels of the transition rate, $q \in \{0.1,\ 0.5,\ 0.9\}$, and three levels of the total demand volume, $D \in \{25,\ 500,\ 2000\}$. Assuming a market share of 3%, these demand volumes translate into realistic customer densities representative of small-sized (0.25

customers/km^2), medium-sized (5 customers/km^2), and large-sized (20 customers/km^2) cities. Finally, we consider three probability vectors representing different degrees of demand imbalance across the consecutive assortment:

$$\mathbf{p} \in \left\{ \left(\frac{1}{2}, \frac{1}{2}\right), \left(\frac{3}{4}, \frac{1}{4}\right), \left(\frac{9}{10}, \frac{1}{10}\right) \right\}.$$

We employ a full factorial design, resulting in $3 \times 3 = 9$ parameterizations of the multinomial distribution and a total of $3 \times 3 \times 9 = 81$ distinct scenarios.

We extend the analysis to consecutive assortments $T_3 \in \mathcal{T}$ consisting of $\hat{n} = 3$ time windows, and their corresponding overlapping assortments $T_5 \in \widetilde{\mathcal{T}}$ comprising $2\hat{n}-1 = 5$ time windows. We evaluate analogous scenarios to gain insights into how the assortment size influences the cost-efficiency of overlapping time windows. Specifically, we retain the same values for the parameters D, z, and q, and consider five probability vectors:

$$\mathbf{p} \in \left\{ \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{4}, \frac{2}{4}, \frac{1}{4}\right), \left(\frac{2}{5}, \frac{1}{5}, \frac{2}{5}\right), \left(\frac{2}{4}, \frac{1}{4}, \frac{1}{4}\right), \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}\right) \right\},$$

resulting in $3 \times 5 = 15$ parameterizations of the multinomial distribution and a total of $3 \times 3 \times 15 = 135$ distinct scenarios.

4.5.2. Numerical Results

We carry out the computational experiments on a Fujitsu Lifebook running Windows 10, equipped with an Intel(R) Core(TM) i5-10310U CPU @ 1.70GHz and four cores. We implement the Monte Carlo simulation in Python 3.13.5 and we solve instances of the linear optimization problem defined in Equation (4.9) using the commercial solver Gurobi Optimizer 12.0.3. All optimization problems are solved to optimality.

For all experiments, we choose the Monte Carlo sample size m such that differences of ± 1 in the estimated expected relative difference $\widehat{\Delta}_{W(\hat{n})}$ are statistically significant at the 95% confidence level. To reduce simulation variability, we use the same set of random samples, drawn from Multinomial$(D, \mathbf{p})$, across all combinations of z and q.

Impact of Demand Distribution, Volume, and Service-to-Routing Ratio on Overlap Performance

In this section, we test the claims stated in Conjecture 4.1, focusing on the scenario where an overlapping time window is added to $\hat{n} = 2$ consecutive time windows. We consider the aggregated average outcomes by parameter setting, summarized in Table 4.2, and the individual interaction effects presented in Figure 4.9, which shows the results of the full factorial analysis. This twofold analysis helps isolate and better understand the individual influence of the three key factors discussed in Conjecture 4.1: (i) the degree of demand imbalance in the consecutive assortment, characterized by the distribution vector $\mathbf{p}$, (ii) the total demand volume D, and (iii) the service-to-routing time ratio z.

Table 4.2.: Aggregated summary of parameters $\mathbf{p}$, D, and z for $\hat{n} = 2$ consecutive time windows

Parameter	**Success Rate**[1]	**Avg.** $\widehat{\Delta}_{W(2)}$
$\mathbf{p} = (0.5, 0.5)$	18.52	27.71
$\mathbf{p} = (0.75, 0.25)$	51.85	4.46
$\mathbf{p} = (0.9, 0.1)$	51.85	-4.77
$D = 25$	29.62	13.64
$D = 500$	44.44	7.82
$D = 2000$	48.15	5.92
$z = 0.1$	22.22	17.88
$z = 0.7$	37.04	5.90
$z = 1.5$	62.96	3.61

[1] Percentage of scenarios where the overlapping assortment led to a statistically significant cost reduction compared to the consecutive assortment.

The success rates reported in Table 4.2 show that overlapping assortments outperform consecutive ones more frequently when demand is unevenly distributed across the consecutive time windows. For the imbalanced settings $\mathbf{p} = (0.75, 0.25)$ and $\mathbf{p} = (0.9, 0.1)$, cost savings are achieved in over 50% of the evaluated scenarios, whereas for the balanced case $\mathbf{p} = (0.5, 0.5)$, the success rate drops significantly to 18.52%. A similar pattern emerges for the other parameters: the success rate increases monotonically with both the total demand volume D and the service-to-routing time ratio z. These findings support the claims made in Conjecture 4.1, which posit that pronounced demand imbalances,

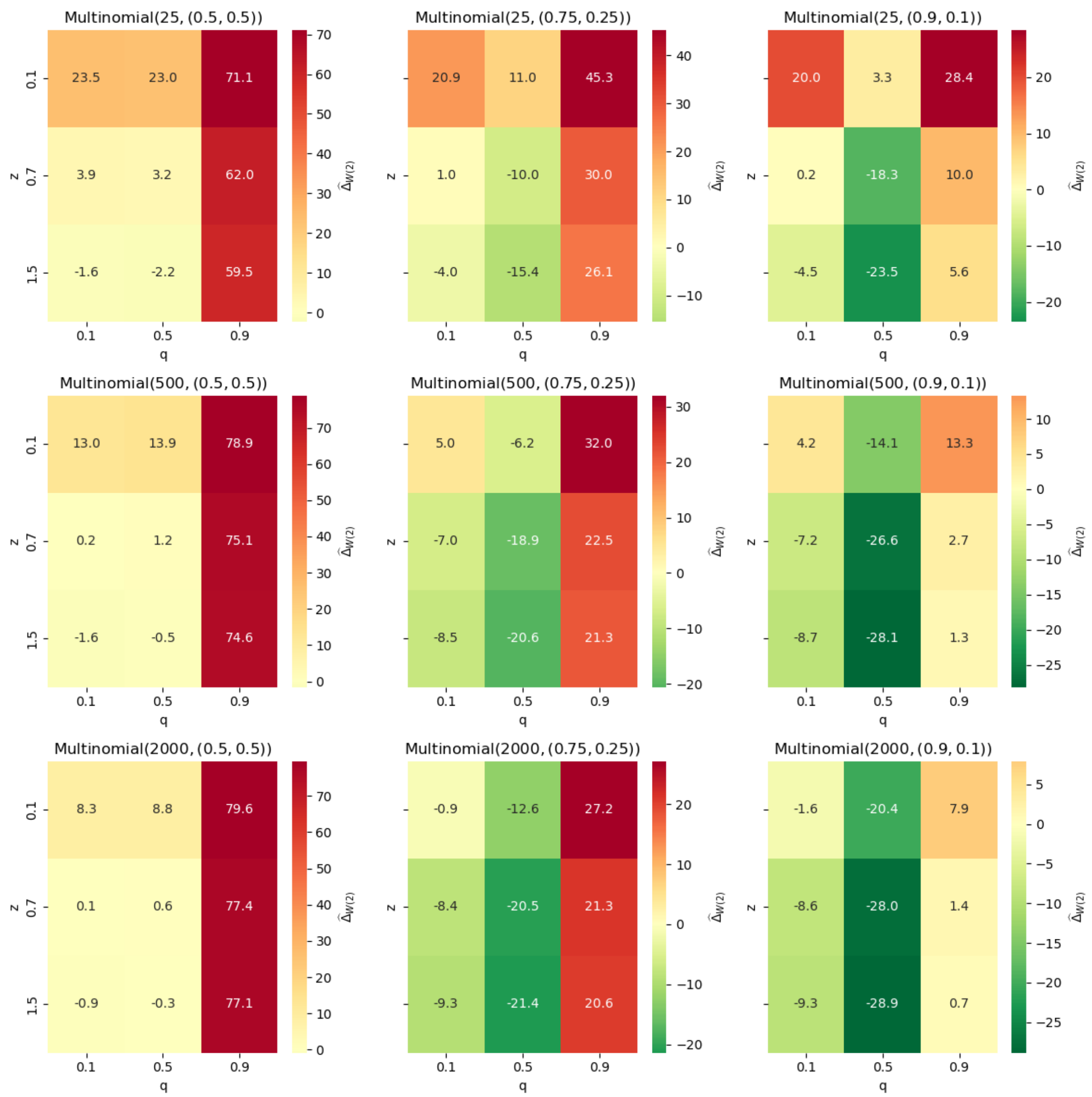


Figure 4.9.: Monte Carlo simulation results for $\hat{n} = 2$ consecutive time windows

high total demand, and a service time that dominates routing time all contribute to potential cost-savings when adopting overlapping time windows.

The aggregated results in Table 4.2 reveal a clear monotonic relationship between each of the three assessed parameters and both the success rate and the estimated expected relative difference of bottleneck workloads $\widehat{\Delta}_{W(2)}$, indicating a positive correlation in all cases. For the probability vector $\mathbf{p}$ and the service-to-routing time ratio z, this relationship also holds consistently at the level of individual scenarios, as shown in Figure 4.9. In these cases, every instance exhibits a performance improvement as the corresponding parameter value increases.

The consistent positive impact of demand imbalance on the relative cost efficiency of overlapping time windows is intuitive: overlapping windows introduce flexibility that allows for reassigning customers to adjacent time windows. This flexibility helps to smoothen out peaks in demand, leading to better-balanced delivery tours and fewer tours required to meet all demand. As a result, the more imbalanced the initial demand across time windows, the greater the potential benefit from overlap-induced reallocation. Interestingly, the success rate in Table 4.2 for balanced demand $\mathbf{p} = (0.5, 0.5)$ shows that overlaps can improve cost performance in 18.52% of the evaluated cases. This suggests that overlaps are not only useful for mitigating structural imbalances (e.g., in expected demand), but also for buffering statistical fluctuations around the expectation.

The effect of the service-to-routing time ratio z is less straightforward. One explanation might be the differing marginal behaviors of service and routing times. Service time grows linearly with the number of customers, whereas routing time exhibits decreasing marginal costs. When routing time dominates, the negative impact of demand peaks is softened by dense customer clustering, which reduces routing effort per stop. Conversely, when service time dominates, concentrated demand peaks increase overall workload without much relief from routing efficiencies. In these cases, smoothing demand via overlapping time windows becomes especially beneficial, as it helps evenly distribute service effort and reduce total costs.

In contrast, Figure 4.9 shows that the impact of the total demand volume D is not consistent across all parameter settings. While a higher D tends to enhance cost savings in scenarios with initial demand imbalances, this pattern does not hold under balanced demand when the expected number of transitions to overlapping options is high. In such cases, increasing D can actually worsen the performance of overlapping time windows. A possible explanation is that increasing D raises the absolute number of customer transitions to overlapping time windows. This can create new, more pronounced bottlenecks, thereby worsening the performance of overlapping assortments. In contrast, the performance of consecutive assortments may continue to improve with higher customer density. As a result, the relative cost difference between the two assortment types increases with growing demand volume. Thus, while higher demand volumes can enhance the cost-efficiency of overlapping assortments, this effect is sensitive to the structure of the underlying demand distribution and customer behavior.

Impact of Demand Transitions on Overlap Performance

This section assesses the impact of the transition rate q, which determines the proportion of customers who shift from a time window in the original consecutive assortment to an added overlapping time window. Together with the initial demand distribution, the transition rate shapes how demand spreads across the overlapping assortment. Specifically, the transition rate determines how the bottleneck demand $\hat{N}$ in the overlapping assortment compares to the bottleneck demand $\bar{N}$ in the consecutive assortment, a relationship that is central to Corollary 4.2, and ultimately influences the cost-efficiency of the overlapping design relative to its consecutive counterpart.

Table 4.3.: Aggregated summary of parameter q for $\hat{n} = 2$ consecutive time windows

Parameter	**Success Rate[1]**	**Avg.** $\widehat{\Delta}_{W(2)}$
$q = 0.1$	55.56	0.67
$q = 0.5$	66.67	-9.31
$q = 0.9$	0.00	36.03

[1] Percentage of scenarios where the overlapping assortment led to a statistically significant cost reduction compared to the consecutive assortment.

The aggregated results in Table 4.3 reveal a non-monotonic relationship between the transition rate q and both the success rate and the estimated expected relative difference of bottleneck workloads $\widehat{\Delta}_{W(2)}$. At low transition rates, expected delivery costs show little difference between the assortments. However, when customers transition more evenly, the overlapping assortment performs significantly better. In contrast, when the additional overlapping option attracts most of the demand, performance deteriorates sharply: there is not a single scenario where overlaps improve cost efficiency in that case.

The success rates can be explained analytically based on how we model customer transitions. For a consecutive assortment T_2 and an overlapping assortment T_3, the bottleneck demand of the overlapping assortment is given by $\hat{N} = qD$. According to Corollary 4.2, a necessary condition for the overlapping assortment to reduce costs is that $\hat{N} < \bar{N}$. However, across all our settings, the expected bottleneck demand in the consecutive assortment satisfies $\bar{N} \leq 0.9$. Therefore, when the transition rate is $q = 0.9$, the overlapping time window becomes the new bottleneck with $\hat{N} = 0.9D$, meaning the necessary condition $\hat{N} < \bar{N}$ is not satisfied in expectation. This explains the observed success rate of 0%.

In contrast, when the transition rate is low at $q = 0.1$, the overlapping time window receives only $0.1D$ of demand, so the bottleneck remains in one of the consecutive time windows at $\hat{N} = 0.9\bar{N}$. This satisfies the necessary condition $\hat{N} < \bar{N}$. Whether the necessary and sufficient condition in Theorem 4.1 is also met then depends on the values of the remaining parameters. For a balanced transition rate of $q = 0.5$, the overlapping time window becomes the new bottleneck with $\hat{N} = 0.5D$. This still satisfies the necessary condition in almost all cases, except when demand over the consecutive assortment is perfectly balanced, that is, when $\bar{N} = 0.5D$.

Furthermore, Figure 4.9 suggests that a bottleneck at $\hat{N} = 0.5D$ (as in the case of $q = 0.5$) is more likely to satisfy the condition in Theorem 4.1 than a bottleneck at $\hat{N} = 0.9\bar{N}$ (as in the case of $q = 0.1$). This implies that more balanced transitions tend to facilitate better smoothing of demand imbalances. Only in cases of initially balanced demand does a transition rate of $q = 0.5$ perform similarly or slightly worse than $q = 0.1$, likely because the demand is already evenly distributed and small fluctuations can be effectively absorbed by minor demand transitions.

Impact of Assortment Size on Overlap Performance

While neither of the conditions in Theorem 4.1 and Corollary 4.2 directly depends on the number of time windows, Lemma 4.2 reveals an explicit dependency in the best- and worst-case performance bounds. Specifically, the gap between performance bounds widens as the best-case performance improves with increasing n, albeit at a decreasing rate, indicating that increased demand variability in larger assortments can positively impact expected delivery costs. To further explore this observation, we examine the impact of assortment size, defined as the number of offered time windows, by comparing earlier results for $T_2 \in \mathcal{T}$ and $T_3 \in \widetilde{\mathcal{T}}$, which consist of $\hat{n} = 2$ consecutive time windows, with new results for $T_3 \in \mathcal{T}$ and $T_5 \in \widetilde{\mathcal{T}}$, which consist of $\hat{n} = 3$ consecutive time windows.

Table 4.4.: Comparative aggregated summary of parameters

Parameter		**Success Rate**[1]		**Avg.** $\widehat{\Delta}_{W(\hat{n})}$	
$\hat{\mathbf{n}} = \mathbf{2}$	$\hat{\mathbf{n}} = \mathbf{3}$	$\hat{\mathbf{n}} = \mathbf{2}$	$\hat{\mathbf{n}} = \mathbf{3}$	$\hat{\mathbf{n}} = \mathbf{2}$	$\hat{\mathbf{n}} = \mathbf{3}$
$\mathbf{p} = (0.5,\ 0.5)$	$\mathbf{p} = (0.33,\ 0.33,\ 0.33)$	18.52	44.44	27.71	13.05
$\mathbf{p} = (0.75,\ 0.25)$	—	51.85	—	4.46	—
$\mathbf{p} = (0.9,\ 0.1)$	—	51.85	—	-4.77	—
—	$\mathbf{p} = (0.25,\ 0.5,\ 0.25)$	—	77.78	—	-19.20
—	$\mathbf{p} = (0.4,\ 0.2,\ 0.4)$	—	40.74	—	6.57
—	$\mathbf{p} = (0.5,\ 0.25,\ 0.25)$	—	48.15	—	2.29
—	$\mathbf{p} = (0.2,\ 0.4,\ 0.4)$	—	48.15	—	7.41
$D = 25$		29.62	42.22	13.64	6.67
$D = 500$		44.44	55.56	7.82	0.62
$D = 2000$		48.15	57.78	5.92	-1.20
$z = 0.1$		22.22	17.78	17.88	12.21
$z = 0.7$		37.04	66.67	5.90	-1.64
$z = 1.5$		62.96	71.11	3.61	-4.50
$q = 0.1$		55.56	68.89	0.67	-7.66
$q = 0.5$		66.67	77.78	-9.31	-9.98
$q = 0.9$		0.00	8.89	36.03	23.71

[1] Percentage of scenarios where the overlapping assortment led to a statistically significant cost reduction compared to the consecutive assortment.

Table 4.4 compares the impact of adding overlapping time windows to assortments with $\hat{n} = 2$ and $\hat{n} = 3$ consecutive time windows, based on the parameter settings introduced in Section 4.5.1. Note that the probability vector $\mathbf{p} = (p_1, p_2, \ldots, p_{\hat{n}})$ is not directly comparable across assortment sizes, as it depends on $\hat{n}$. This dependency is carefully accounted for in all interpretations and explicitly indicated in Table 4.4: each row represents a pair of comparable parameter settings for the assortment sizes $\hat{n} = 2$ and $\hat{n} = 3$. A dash denotes the absence of a comparable setting, while blank cells indicate identical settings. Despite these differences, the selected scenarios ensure comparable coverage of the parameter space across both assortment sizes.

First, we observe that the aggregate-level relationships involving the parameters D, z, and q remain consistent across both the smaller and larger assortments. Moreover, the larger assortment consistently outperforms the smaller one at the aggregated parameter level across all configurations, except in the case where $z = 0.1$. For this configuration,

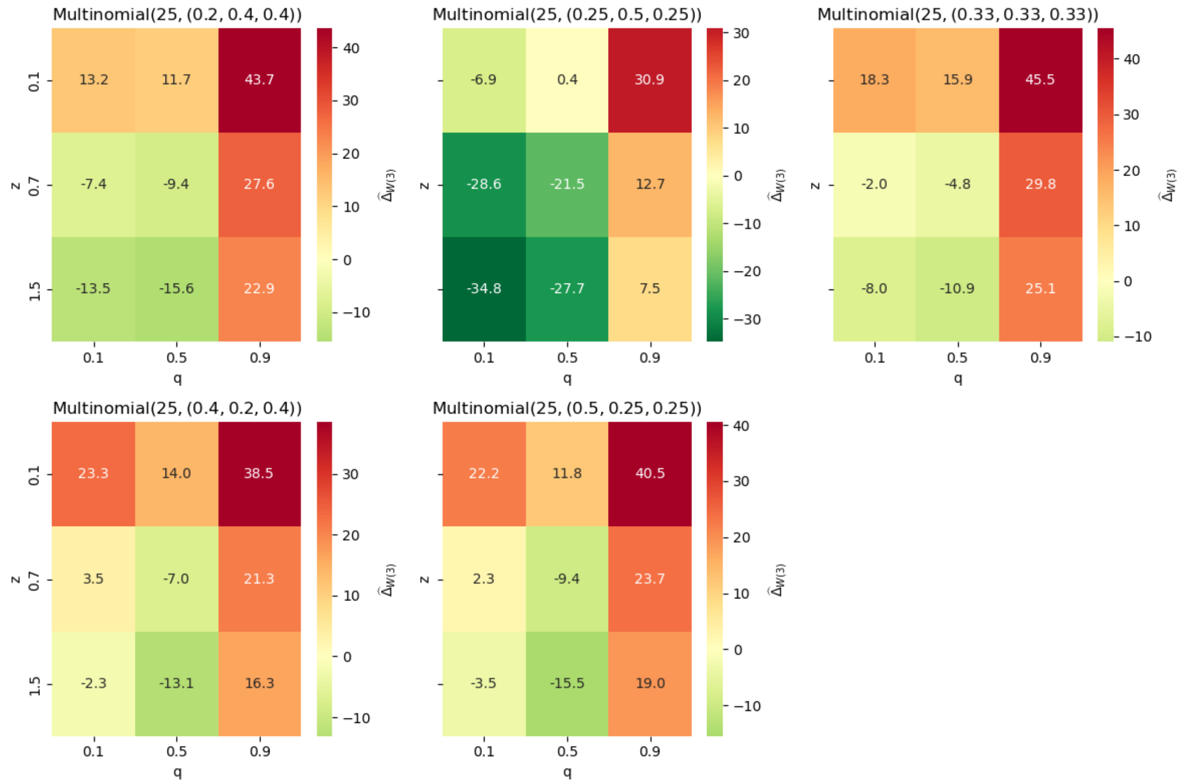


Figure 4.10.: Monte Carlo simulation results for $\hat{n} = 3$ consecutive time windows and $D = 25$

the success rate is 22.22% for the small assortment but only 17.78% for the large assortment. However, since the number of scenarios differs between the two assortments, it is also helpful to consider absolute counts: the large assortment achieves a cost reduction in 8 out of 45 scenarios, while the small assortment does so in only 6 out of 27 (see Appendix C for the full table of results).

The average relative difference $\widehat{\Delta}_{W(\cdot)}$ further supports this interpretation, decreasing from a 17.88% higher bottleneck workload when overlaps are added to the small assortment to only 12.21% in the case of the larger assortment. These results suggest that the observed deviation for $z = 0.1$ does not indicate a structural effect associated with low service-to-routing time ratios, but is more likely attributable to the specific composition of the distribution vector $\mathbf{p}$. Setting aside this point, we conclude that assortment size generally exerts a positive effect on the cost-efficiency of overlapping time windows.

Lastly, we assess the results based on the distribution vector $\mathbf{p}$. The only vector setting comparable between the small and large assortments is the case of expected balanced demand. In this case, Table 4.4 shows that the larger assortment again outperforms the smaller one: the success rate increases from 18.52% to 44.44%, indicating a cost improvement in nearly half of the cases, while the average relative performance difference decreases from a 27.71% higher bottleneck workload to only 13.05%.

When assessing the performance of the large assortments across all considered configurations of $\mathbf{p}$, one might expect overlapping time windows to perform worst under the balanced vector $\mathbf{p} = (0.33, 0.33, 0.33)$, as hypothesized in Conjecture 4.1. Interestingly, however, the imbalanced probability vector $\mathbf{p} = (0.4, 0.2, 0.4)$ has a slightly lower success rate (11 out of 27) compared to the balanced vector (12 out of 27). However, it outperforms the balanced case in terms of average relative difference in bottleneck workloads: 6.57% versus 13.05%, respectively.

To explain this deviation, we analyze individual interaction effects presented in Figure 4.10, an excerpt from the Monte Carlo results for a small demand volume of $D = 25$. We find that the imbalanced vector underperforms the balanced one when the transition rate is low ($q = 0.1$), accounting for the deviation noted above. One explanation for this observation is that the imbalanced vector leads to demand distributions in which fewer customers choose the middle time window, creating peaks at the beginning and end of the delivery shift. When overlaps are introduced but only a few customers transition, these peaks are amplified, resulting in worse performance than in the balanced case, where overlaps help smoothen statistical demand fluctuations. This finding underscores the importance of jointly considering the initial demand distribution and expected customer transitions when assessing the value of overlapping time windows.

4.6. Conclusion

This paper advances the understanding of time window assortment design for attended home delivery by focusing on the role of overlapping time windows under demand uncertainty. Building on previous work that introduced time window assortment design as a planning problem, the specific value of time window overlaps in mitigating operational inefficiencies caused by demand fluctuations remained unclear. To address this gap, we extend an established assortment evaluation model to incorporate demand variability and develop analytical and simulation-based results that shed light on the cost implications of offering overlapping time windows.

We derive necessary and sufficient conditions under which demand realizations for overlapping time windows lead to lower total delivery costs compared to consecutive designs. Since demand is still uncertain at the time of the assortment decision, we use the theoretical results to form hypotheses about when overlaps are likely to reduce costs. To test these hypotheses and explore the interaction effects between key parameters, we conduct a Monte Carlo simulation that quantifies the expected performance gap between overlapping and consecutive assortments.

Our results show that overlapping time windows offer the flexibility to effectively smooth operational bottlenecks by reallocating workload across adjacent time windows, but only under certain conditions. Their effectiveness depends on the interplay between operational parameters, demand characteristics, and customer behavior. Overlaps are particularly beneficial when service time dominates routing time and total demand is high. However, their success hinges on the relationship between transition rates and the initial demand distribution: in some cases, the added flexibility is underutilized and fails to sufficiently smooth operational bottlenecks, falling short of offsetting the increased constraint density introduced by the overlaps; in others, new bottlenecks arise, unavoidably driving up overall delivery costs. Furthermore, while initial demand imbalances can amplify the benefits of overlaps, this effect is not universal and can be offset by unfavorable transition dynamics. Lastly, larger assortments tend to benefit more from overlaps due to greater demand variability, provided demand remains sufficiently dispersed. Overall, overlaps should be viewed not as a default feature but as a strategic design choice, whose value strongly depends on expected demand patterns and customer choice behavior.

Our analysis relies on several simplifying assumptions that frame the interpretation of the results. Specifically, we focus exclusively on the cost perspective, assessing the operational efficiency of overlapping time windows without considering their potential to attract additional demand (Waßmuth et al., 2025). Additionally, customer transitions between time windows are modeled deterministically through an exogenous parameter q, which limits the behavioral complexity captured by the model. Consequently, while our findings provide valuable insights into the structural factors affecting the cost efficiency of overlapping time windows, they may not fully capture all possible interaction effects simultaneously.

Future research could jointly assess the cost and revenue implications of overlapping time windows and endogenize customer choice behavior when transitioning to overlapping time window, allowing for a richer analysis of interaction effects. Another promising avenue lies in the joint optimization of time window assortment design and capacity planning. Finally, empirical validation using operational data could help test the model's assumptions and assess the practical relevance of the findings in real-world delivery environments.

Appendix A

Proofs of Chapter III

Proof of Lemma 3.1. First, we introduce the index $i = 1, \ldots, n$ to distinguish time windows in a given overlapping time window assortment $T \in \widetilde{\mathcal{T}}$. Consider allocating a given demand $\mathbf{N} = (N_i)_{i=1,\ldots,n}$ from n overlapping time windows of length l to $(n+1)$ consecutive delivery intervals of length $\frac{l}{2}$. We model demand allocation using weights $\boldsymbol{\eta} = (\eta_i \in [0,1])_{i=1,\ldots,n}$ which represent the fraction of demand allocated to the first half of the corresponding time window. Accordingly, we define allocation functions $\boldsymbol{\theta} = (\theta_i(\boldsymbol{\eta}, \mathbf{N}))_{i=1,\ldots,n+1}$ that determine the demand in the consecutive delivery intervals. Second, the delivery costs $\widetilde{C}(T, \boldsymbol{\theta}) = \widetilde{C}^t(T) \cdot \widetilde{v}(T, \boldsymbol{\theta})$ depend on the allocated demand only for the number of tours. Given that vehicle tours are homogeneous (Assumption 3.2), the number of vehicle tours becomes

$$\widetilde{v}(\boldsymbol{\theta}) = \frac{2}{l} \max_{i=1,\ldots,n+1} \left(\tau\theta_i(\boldsymbol{\eta}, \mathbf{N}) + \alpha k \sqrt{R\theta_i(\boldsymbol{\eta}, \mathbf{N})} \right)$$

Consequently, to minimize the delivery costs after demand allocation, we seek the weights (i.e., the demand allocation) that minimize the largest workload per delivery interval, resulting in the following nonlinear optimization model:

$$\begin{aligned}
\min_{\boldsymbol{\eta}} \quad & \max_{i=1,\ldots,n+1} \left(\tau\theta_i(\boldsymbol{\eta}, \mathbf{N}) + \alpha k \sqrt{R\theta_i(\boldsymbol{\eta}, \mathbf{N})} \right) & \\
\text{s.t.} \quad & \theta_i(\boldsymbol{\eta}, \mathbf{N}) = \eta_i N_i & i = 1, \\
& \theta_i(\boldsymbol{\eta}, \mathbf{N}) = (1 - \eta_{i-1}) N_{i-1} + \eta_i N_i & i = 2, \ldots, n, \\
& \theta_i(\boldsymbol{\eta}, \mathbf{N}) = (1 - \eta_{i-1}) N_{i-1} & i = n+1, \\
& \eta_i \in [0,1] & i = 1, \ldots, n
\end{aligned}$$

Obviously, the demand allocation that minimizes the largest workload per delivery interval is $\boldsymbol{\theta}^* = \mathbf{1}_{n+1} \cdot \left(\frac{1}{n+1}\sum_{i=1}^{n} N_i\right)$, meaning that an even distribution of demand across consecutive delivery intervals is an optimal solution to the above problem. Given homogeneous demand (Assumption 3.2), the initial demand is evenly distributed over the assortment, so there exist optimal weights $\boldsymbol{\eta}^*$ such that this solution is feasible.

□

Proof of Lemma 3.2. Consider Equation (3.1) for a given demand density $\delta = \frac{N}{R} > 0$ and a maximum vehicle capacity $\bar{Q} > 0$. For time window assortments $T \in \mathcal{T}$, we get:
Case 1: The maximum vehicle capacity is more constraining than the (time window-driven) required vehicle capacity. Formally, this means:

$$\begin{aligned} & \bar{Q} < Q(T, N) = \frac{L(T)}{w^c(N)} = \frac{nl}{\tau + \frac{\alpha k}{\sqrt{\delta}}} \\ \Leftrightarrow \quad & \frac{\bar{Q}}{n} < \frac{l}{\tau + \frac{\alpha k}{\sqrt{\delta}}} \end{aligned}$$

Case 2: The maximum vehicle capacity is more constraining than the total demand. Formally, this means:

$$\begin{aligned} & \bar{Q} < nN = n\delta R \\ \leftrightarrow \quad & \frac{\bar{Q}}{n} < \delta R \end{aligned}$$

□

Proof of Theorem 3.1. Consider Equations (3.1) and (3.2) and a maximum vehicle capacity $\bar{Q} > 0$, and let the net margin per order be positive, $r > c\tau$. For time window assortments $T \in \mathcal{T}$, we get:

We first show that the profit function is convex in the time window demand $N > 0$:

$$\begin{aligned} P(T,N) &= rnN - \frac{C^t(T)}{l}(\tau N + \alpha k\sqrt{RN}) \\ \frac{\partial P}{\partial N}(T,N) &= nr - \frac{C^t(T)}{l}\left(\tau + \frac{1}{2}\alpha k\sqrt{\frac{R}{N}}\right) \\ \frac{\partial^2 P}{\partial N^2}(T,N) &= \frac{C^t(T)}{4l}\alpha k\sqrt{\frac{R}{N^3}} > 0 \end{aligned}$$

We then derive the conditions on the delivery shift length $L(T)$ that lead to the performance regions of a) unprofitable, b) multi-tour profit, and c) single-tour profit. First, we show that for small $L(T)$, the profit function is convex decreasing $\forall N > 0$:

$$\begin{aligned} & \frac{\partial P}{\partial N}(T,N) \overset{!}{\leq} 0,\ \forall N > 0 \\ \Leftrightarrow\quad & rn - \frac{C^t(T)}{l}\left(\tau + \frac{1}{2}\alpha k\sqrt{\frac{R}{N}}\right) \leq 0,\ \forall N > 0 \\ \Leftrightarrow\quad & rL(T) \leq C^t(T)\left(\tau + \frac{1}{2}\alpha k\sqrt{\frac{R}{N}}\right),\ \forall N > 0 \\ \Leftrightarrow\quad & rL(T) - C^t(T)\tau \leq C^t(T)\left(\frac{1}{2}\alpha k\sqrt{\frac{R}{N}}\right),\ \forall N > 0 \\ \Leftrightarrow\quad & rL(T) - C^t(T)\tau \leq 0 \\ \Leftrightarrow\quad & rL(T) - (cL + F)\tau \leq 0 \\ \Leftrightarrow\quad & L(T) \leq \frac{F\tau}{r - c\tau} \end{aligned}$$

Conversely, we know that for $L(T) > \frac{F\tau}{r-c\tau}$, the exists a global minimum at a $N > 0$ after which $P(T,N)$ is increasing. Next, we need to ensure that $v(T,N) \geq 1$. We break even for exactly one complete vehicle tour, given a demand N^*, if and only if the profit contribution of N^* equals the fixed cost per tour:

$$
\begin{aligned}
& (r - cw^c(N^*))\, nN^* \stackrel{!}{=} F \\
\Leftrightarrow \quad & (r - c\tau)nN^* - cn\alpha k\sqrt{RN^*} = F \\
\Leftrightarrow \quad & (N^*)^2 - \frac{2(r-c\tau)F + n(c\alpha k)^2 R}{n(r-c\tau)^2} N^* + \left(\frac{F}{n(r-c\tau)}\right)^2 = 0 \\
\Leftrightarrow \quad & N^*(n) := \frac{2(r-c\tau)F + n(c\alpha k)^2 R}{2n(r-c\tau)^2} + \sqrt{\left(\frac{2(r-c\tau)F + n(c\alpha k)^2 R}{2n(r-c\tau)^2}\right)^2 - \left(\frac{F}{n(r-c\tau)}\right)^2}
\end{aligned}
$$

The break-even shift length $L^P(N)$ for any demand $N > 0$ is

$$
\begin{aligned}
& P(T, N) \stackrel{!}{=} 0 \\
\Leftrightarrow \quad & (r - cw^c(N))\, nN - F\frac{w(N)}{l} = 0 \\
\Leftrightarrow \quad & l = \frac{Fw(N)}{(r - cw^c(N))\, nN} \\
\Leftrightarrow \quad & nl = \frac{Fw(N)}{(r - cw^c(N))\, N} \\
\Leftrightarrow \quad & L^P(N) := \frac{Fw^c(N)}{r - cw^c(N)}
\end{aligned}
$$

Consequently, for $L(T) \geq \frac{Fw^c(N^*(n))}{r - cw^c(N^*(n))}$, the delivery system needs to prioritize utilizing one complete vehicle tour. It holds that $nN^*(n)w\,(N^*n)) = \frac{Fw^c(N^*(n))}{r - cw^c(N^*(n))}$.

Having identified the performance regions defined by $L(T)$, we proceed to characterize the profit behavior within each region:

a) Consider $L(T) \leq \frac{F\tau}{r - c\tau}$: Since $P(T, N)$ is convex decreasing in N and $P(T, 0) = 0$, there exists no demand $N > 0$ such that $P(T, N) \geq 0$.

b) Consider $L(T) \in \left(\frac{F\tau}{r - c\tau}, \frac{Fw^c(N^*(n))}{r - cw^c(N^*(n))}\right)$: Since there exists a global minimum at a $N > 0$ after which $P(T, N)$ is increasing and since $P(T, 0) = 0$, there exists a break-even demand N^P with $P(T, N^P) = 0$ after which the profit function is positive and increasing. We solve for N^P:

$$
\begin{aligned}
& P(T, N^P) \stackrel{!}{=} 0 \\
\Leftrightarrow \quad & rnN^P - \frac{C^t(T)}{l}(\tau N^P + \alpha k\sqrt{RN^P}) = 0 \\
\Leftrightarrow \quad & rL(T)N^P - C^t(T)\tau N^P - C^t(T)\alpha k\sqrt{RN^P} = 0 \\
\Leftrightarrow \quad & N^P(rL(T) - C^t(T)\tau) = C^t(T)\alpha k\sqrt{RN^P} \\
\Leftrightarrow \quad & \sqrt{N^P} = \frac{C^t(T)\alpha k\sqrt{R}}{rL(T) - C^t(T)\tau} \\
\Leftrightarrow \quad & N^P(T) = \left(\frac{C^t(T)\alpha k\sqrt{R}}{rL(T) - C^t(T)\tau}\right)^2 > 0 \\
\Leftrightarrow \quad & \delta^P(T) := \frac{N^P(T)}{R} = \left(\frac{C^t(T)\alpha k}{rL(T) - C^t(T)\tau}\right)^2
\end{aligned}
$$

with $rL(T) - C^t(T)\tau > 0$ since $L(T) > \frac{F\tau}{r - c\tau}$.

Second, we assess the behavior of the required vehicle capacity in $N > 0$ to estimate the required capacity if we were to increase demand beyond the break-even point:

$$
\begin{aligned}
Q(T, N) &= \frac{L(T)}{w^c(N)} \\
\frac{\partial Q}{\partial N}(T, N) &= \frac{L(T)\alpha k\sqrt{\frac{R}{N}}}{2w(N)} > 0
\end{aligned}
$$

Thus, the required vehicle capacity is strictly increasing in $N > 0$. The vehicle capacity required to serve demand density $\delta^P(T)$ is $Q(T, \delta^P) = \frac{L(T)}{\tau + \frac{\alpha k}{\delta^P(T)}}$. Therefore, for any density $\delta \geq \delta^P(T)$ and any shift length $L(T)$ within the specified range, the required vehicle capacity satisfies $Q(T, \delta) \geq \frac{L(T)}{\tau + \frac{\alpha k}{\sqrt{\delta^P(T)}}}$.

c) Consider $L \geq \frac{Fw^c(N^*(n))}{r - cw^c(N^*(n))}$: We determine the minimum demand that ensures at least one complete vehicle tour. So we look for a demand N^v with a workload equal to the time window length implied by the shift length $L(T) = nl$:

$$\begin{aligned} & l \overset{!}{=} \tau N^v + \alpha k \sqrt{R N^v} \\ \Leftrightarrow \quad & N^v(l) = \frac{2l\tau + (\alpha k)^2 R - \alpha k \sqrt{R} \sqrt{4l\tau + (\alpha k)^2 R}}{2\tau^2} \\ \Leftrightarrow \quad & \delta^v(l) := \frac{N^v(l)}{R} = \frac{2l\tau + (\alpha k)^2 R - \sqrt{(2l\tau + (\alpha k)^2 R)^2 - (2l\tau)^2}}{2R\tau^2} \end{aligned}$$

Since the break-even demand decreases in $L(T)$, we have $N^v(l) \geq N^P(T) > 0$. As profits increase in $N \geq N^P(T)$, this property implies non-negative profits, $P(T, N^v(l)) \geq 0$.

Second, we estimate the required capacity if we were to increase demand beyond the tour-utilization point. The vehicle capacity required to serve demand density $\delta^v(l)$ is $Q(T, \delta^v) = n\delta^v(l)R$. Since the required vehicle capacity is strictly increasing in $N > 0$, it holds that for any density $\delta \geq \delta^v(l)$ and any shift length $L(T)$ within the specified range, the required vehicle capacity satisfies $Q(T, \delta) \geq n\delta^v(l)R$.

□

Proof of Lemma 3.3. Consider Equations (3.1) and (3.2) for a given number of $n \geq 1$ consecutive time windows and demand per time window $N(l)$ that is independent of the time window length, $\frac{\partial N}{\partial l}(l) = 0$. For time window lengths $l > 0$ and demand $N(l) > 0$ we get:

a) $\frac{\partial w^c}{\partial l}(l) = 0$ and $\frac{\partial w}{\partial l}(l) = 0$

b) $\frac{\partial v}{\partial l}(l) = -\frac{w(l)}{l^2} < 0$

c) $\frac{\partial P}{\partial l}(l) = \frac{F w(l)}{l^2} > 0$

d) $\frac{\partial Q}{\partial l}(l) = \frac{n}{w^c(l)} > 0$

□

Proof of Proposition 3.1. Consider Equations (3.1) and (3.2) for a given number of $n \geq 1$ consecutive time windows and demand per time window $N(l) = \hat{N} > 0$. For time windows of length $l < w(l)$ and sufficiently large maximum vehicle capacity $Q(l) \leq \bar{Q}$, we get: From Lemma 3.3 we know that profit and the required vehicle capacity increase in l.

□

Proof of Lemma 3.4. Consider Equations (3.1) and (3.2) for a given number of $n \geq 1$ consecutive time windows and demand per time window $N(l)$ that is decreasing in the time window length, $\frac{\partial N}{\partial l}(l) < 0$. For time window lengths $l > 0$ and demand $N(l) > 0$ we get:

a) $\frac{\partial w^c}{\partial l}(l) = -\frac{1}{2}\alpha k\sqrt{\frac{R}{N(l)^3}}\frac{\partial N}{\partial l}(l) > 0$

b) $\frac{\partial w}{\partial l}(l) = \left(\tau + \frac{1}{2}\alpha k\sqrt{\frac{R}{N(l)}}\right)\frac{\partial N}{\partial l}(l) < 0$

c) $\frac{\partial v}{\partial l}(l) = \frac{1}{l}\frac{\partial w}{\partial l}(l) - \frac{1}{l^2}w(l) < 0$

d) $P(l) = rnN(l) - cnw(l) - Fv(l)$

$$\begin{aligned}\frac{\partial P}{\partial l}(l) &= rn\frac{\partial N}{\partial l}(l) - cn\frac{\partial w}{\partial l}(l) - F\frac{\partial v}{\partial l}(l)\\ &= rn\frac{\partial N}{\partial l}(l) - cn\left(\tau + \frac{1}{2}\alpha k\sqrt{\frac{R}{N(l)}}\right)\frac{\partial N}{\partial l}(l) - F\frac{\partial v}{\partial l}(l)\\ &= \left(n(r - c\tau) - \frac{1}{2}nc\alpha k\sqrt{\frac{R}{N(l)}}\right)\frac{\partial N}{\partial l}(l) - F\frac{\partial v}{\partial l}(l)\end{aligned}$$

e) $Q(l) = \frac{nN(l)}{v(l)} = \frac{nl}{w^c(l)}$

$$\begin{aligned}\frac{\partial Q}{\partial l}(l) &= \frac{n}{v(l)}\left(\frac{\partial N}{\partial l}(l) - \frac{N(l)}{v(l)}\frac{\partial v}{\partial l}(l)\right)\\ &= \frac{n}{w^c(l)}\left(1 - \frac{l}{w^c(l)}\frac{\partial w^c}{\partial l}(l)\right)\end{aligned}$$

□

Proof of Proposition 3.2. Consider Equations (3.1) and (3.2) for a given number of $n \geq 1$ consecutive time windows and demand $N(l) = \bar{N} - \gamma l$ with a demand potential of $\bar{N} > 0$ per time window and a time sensitivity factor $\gamma > 0$. Let the net margin per order be positive, $r > c\tau$. For time windows of length $l < w(l)$, meaning $l < \frac{\tau\bar{N}+\gamma\tau^2\bar{N}-\frac{\gamma}{2}(\alpha k)^2 R+\frac{1}{2}\alpha k\sqrt{R}\sqrt{(4\gamma\tau\bar{N}+4\bar{N}+(\gamma\alpha k)^2 R)}}{\gamma^2\tau^2+2\gamma\tau+1}$, and sufficiently large maximum vehicle capacity $Q(l) \leq \bar{Q}$, we get: Increasing the time window length l increases profit if and only if

$$
\begin{aligned}
& \frac{\partial P}{\partial l}(l) \overset{!}{>} 0 \\
\Leftrightarrow \quad & -\gamma n(r - c\tau) + \gamma n \left(\frac{c\alpha k\sqrt{R}}{2\sqrt{N(l)}} \right) - F\frac{\partial v}{\partial l}(l) > 0 \\
\Leftrightarrow \quad & \left(\frac{\gamma n c\alpha k\sqrt{R}}{2\sqrt{N(l)}} \right) > \gamma n(r - c\tau) + F\frac{\partial v}{\partial l}(l) \\
\Leftrightarrow \quad & \gamma n c\alpha k\sqrt{\frac{R}{N(l)}} > 2\left(\gamma n(r - c\tau) + F\frac{\partial v}{\partial l}(l) \right) \\
\Leftrightarrow \quad & \gamma n c\alpha k > 2\left(\gamma n(r - c\tau) + F\frac{\partial v}{\partial l}(l) \right)\sqrt{\frac{N(l)}{R}}
\end{aligned}
$$

<u>Case 1:</u> $\gamma n(r - c\tau) + F\frac{\partial v}{\partial l}(l) \leq 0 \Rightarrow \frac{\partial P}{\partial l}(l) > 0$. The condition can be reformulated as

$$
\begin{aligned}
& \gamma n(r - c\tau) + F\frac{\partial v}{\partial l}(l) \leq 0 \\
\Leftrightarrow \quad & \gamma n(r - c\tau) + \frac{F\gamma\alpha k\sqrt{R}}{2l\sqrt{N(l)}} - \frac{F\tau\bar{N}}{l^2} - \frac{F\bar{N}\alpha k\sqrt{R}}{l^2\sqrt{N(l)}} \leq 0 \\
\Leftrightarrow \quad & l \leq \frac{\sqrt{(F\alpha k)^2\gamma R + 16n\bar{N}(r - c\tau)Fw(l)} - F\alpha k\sqrt{\gamma R}}{4n(r - c\tau)\sqrt{\gamma N(l)}} \\
\Leftrightarrow \quad & l \leq \frac{\sqrt{(F\alpha k)^2\frac{R}{N(l)} + 16n(r - c\tau)\frac{F}{\gamma}\bar{N}w^c(l)} - F\alpha k\sqrt{\frac{R}{N(l)}}}{4n(r - c\tau)} \\
\Leftrightarrow \quad & l \leq \frac{\sqrt{\mathcal{D}(l) + \mathcal{W}(l)} - \sqrt{\mathcal{D}(l)}}{4n(r - c\tau)}
\end{aligned}
$$

with $\mathcal{D}(l) := \frac{(F\alpha k)^2}{\delta(l)}$ and $\mathcal{W}(l) := 16n(r - c\tau)\frac{F}{\gamma}\bar{N}w^c(l)$.

Case 2: $\gamma n(r-c\tau)+F\frac{\partial v}{\partial l}(l)>0 \Leftrightarrow l>\frac{\sqrt{\mathcal{D}(l)+\mathcal{W}(l)}-\sqrt{\mathcal{D}(l)}}{2\mathcal{M}}$. For $\frac{\partial P}{\partial l}(l)>0$ to hold, we need

$$
\begin{aligned}
\Leftrightarrow \quad & \sqrt{\frac{N(l)}{R}} < \frac{\gamma nc\alpha k}{2\left(\gamma n(r-c\tau)+F\frac{\partial v}{\partial l}(l)\right)} \\
\Leftrightarrow \quad & \frac{N(l)}{R} < \left(\frac{\gamma nc\alpha k}{2\left(\gamma n(r-c\tau)+F\frac{\partial v}{\partial l}(l)\right)}\right)^2 \\
\Leftrightarrow \quad & \delta(l) < \left(\frac{\gamma nc\alpha k}{2\mathcal{V}(l)}\right)^2
\end{aligned}
$$

with demand density $\delta(l)=\frac{N(l)}{R}$ and $\mathcal{V}(l):=\gamma n(r-c\tau)+F\frac{\partial v}{\partial l}(l)>0$.

Next, we show that the capacity requirements function increases up to a certain time window length and then decreases. We derive a threshold for the time window length that is a sufficient condition for the vehicle capacity requirements to increase in l.

$$
\begin{aligned}
& \frac{\partial Q}{\partial l}(l) \overset{!}{>} 0 \\
\Leftrightarrow \quad & \frac{n\left(\gamma l\left(2\tau(\bar{N}-\gamma l)+\alpha k\sqrt{R(\bar{N}-\gamma l)}\right)+2(\bar{N}-2\gamma l)\left(\tau(\bar{N}-\gamma l)+\alpha k\sqrt{R(\bar{N}-\gamma l)}\right)\right)}{2\left(\tau(\bar{N}-\gamma l)+\alpha k\sqrt{R(\bar{N}-\gamma l)}\right)^2} > 0 \\
\Leftrightarrow \quad & \gamma l\left(2\tau(\bar{N}-\gamma l)+\alpha k\sqrt{R(\bar{N}-\gamma l)}\right)+2(\bar{N}-2\gamma l)\left(\tau(\bar{N}-\gamma l)+\alpha k\sqrt{R(\bar{N}-\gamma l)}\right) > 0 \\
\Leftrightarrow \quad & 2\bar{N}\tau(\bar{N}-\gamma l)-2\gamma l\tau(\bar{N}-\gamma l)+2\bar{N}\alpha k\sqrt{R(\bar{N}-\gamma l)}-3\gamma l\alpha k\sqrt{R(\bar{N}-\gamma l)} > 0 \\
\Leftrightarrow \quad & (\bar{N}-\gamma l)2\tau(\bar{N}-\gamma l)+(2\bar{N}-3\gamma l)\alpha k\sqrt{R(\bar{N}-\gamma l)} > 0 \\
\Leftrightarrow \quad & 2\tau(\bar{N}-\gamma l)^2+(2\bar{N}-3\gamma l)\alpha k\sqrt{R(\bar{N}-\gamma l)} > 0 \\
\Leftarrow \quad & 2\bar{N}-3\gamma l \geq 0 \\
\Leftrightarrow \quad & l \leq \frac{2}{3}\frac{\bar{N}}{\gamma}
\end{aligned}
$$

□

Proof of Proposition 3.3. Consider Equations (3.1) to (3.4) for a given time window length $l>0$ and demand $N(n)=\frac{\bar{N}}{n}$ with a total demand volume $\bar{N}\geq 2N^v(l)$.

1. We first show that the workload and the number of tours decrease in n:

$$\begin{aligned} v(n+1) &= \frac{w(n+1)}{l} = \frac{\tau\frac{\bar{N}}{n+1} + \alpha k\sqrt{R\frac{\bar{N}}{n+1}}}{l} \\ &= \frac{\tau\bar{N} + \alpha k\sqrt{R(n+1)\bar{N}}}{(n+1)l} = \frac{\tau\bar{N}}{(n+1)l} + \frac{\alpha k\sqrt{R\bar{N}}}{\sqrt{n+1}l} \\ &< \frac{\tau\bar{N}}{nl} + \frac{\alpha k\sqrt{R\bar{N}}}{\sqrt{n}l} = \frac{\tau\bar{N} + \alpha k\sqrt{Rn\bar{N}}}{nl} = \frac{\tau\frac{\bar{N}}{n} + \alpha k\sqrt{R\frac{\bar{N}}{n}}}{l} = \frac{w(n)}{l} \\ &= v(n) \end{aligned}$$

Next, we show that the required vehicle capacity is a strictly increasing function in n:

$$Q(n+1) = \frac{\bar{N}}{v(n+1)} = \frac{v(n)}{v(n+1)}\frac{\bar{N}}{v(n)} = \frac{v(n)}{v(n+1)}Q(n) > Q(n)$$

Now, consider $1 \leq \hat{n} \leq \frac{\bar{N}}{N^v(l)} - 1$ consecutive time windows and sufficiently large maximum vehicle capacity $Q(\hat{n}) \leq \bar{Q}$. We show the necessary and sufficient condition that ensures increasing profits:

$$\begin{aligned} & P(\hat{n}+1) - P(\hat{n}) \overset{!}{>} 0 \\ \Leftrightarrow\quad & \left(\sqrt{\hat{n}} - \sqrt{\hat{n}+1}\right)c\alpha k\sqrt{R\bar{N}} + \left(w(\hat{n}) - w(\hat{n}+1)\right)\frac{F}{l} > 0 \\ \Leftrightarrow\quad & \left(w(\hat{n}) - w(\hat{n}+1)\right)\frac{F}{l} - \left(\sqrt{\hat{n}+1} - \sqrt{\hat{n}}\right)c\alpha k\sqrt{R\bar{N}} > 0 \\ \Leftrightarrow\quad & \left(w(\hat{n}) - w(\hat{n}+1)\right)\frac{F}{l} > \left(\sqrt{\hat{n}+1} - \sqrt{\hat{n}}\right)c\alpha k\sqrt{R\bar{N}} \\ \Leftrightarrow\quad & \sqrt{\bar{N}} > \frac{\alpha k\sqrt{R}}{\tau}(\hat{n}+1)\hat{n}\left(\sqrt{\hat{n}+1} - \sqrt{\hat{n}}\right)\left(\frac{lc}{F} - \frac{1}{\sqrt{\hat{n}+1}\sqrt{\hat{n}}}\right) \\ \Leftrightarrow\quad & \sqrt{\frac{\bar{N}}{R}} > \left(\sqrt{\hat{n}+1} - \sqrt{\hat{n}}\right)\frac{(\hat{n}+1)\hat{n}\alpha k}{\tau}\left(\frac{lc}{F} - \frac{1}{\sqrt{\hat{n}+1}\sqrt{\hat{n}}}\right) \\ \Leftrightarrow\quad & \frac{\bar{N}}{R} > \left(\frac{lc\alpha k(\hat{n}+1)\hat{n}}{F\tau(\sqrt{\hat{n}+1} + \sqrt{\hat{n}})} - \frac{\alpha k\sqrt{(\hat{n}+1)\hat{n}}}{\tau(\sqrt{\hat{n}+1} + \sqrt{\hat{n}})}\right)^2 \\ \Leftrightarrow\quad & \bar{\delta} > \left(\frac{\alpha k\sqrt{(\hat{n}+1)\hat{n}}}{\tau(\sqrt{\hat{n}+1} + \sqrt{\hat{n}})}\right)^2\left(\frac{lc\sqrt{(\hat{n}+1)\hat{n}}}{F} - 1\right)^2 \end{aligned}$$

with total demand density $\bar{\delta} = \frac{\bar{N}}{R}$.

2. Consider $2 \leq \hat{n} \leq \frac{\bar{N}}{N^v(l)}$ consecutive time windows and sufficiently large maximum vehicle capacity $Q(\hat{n}) \leq \bar{Q}$. For the workload, we first show that $\widetilde{w}(2\hat{n}-1) = w(2\hat{n})$:

$$\widetilde{w}(2\hat{n}-1) = \tau \frac{2\hat{n}-1}{2\hat{n}} \frac{\bar{N}}{2\hat{n}-1} + \alpha k \sqrt{R \frac{2\hat{n}-1}{2\hat{n}} \frac{\bar{N}}{2\hat{n}-1}} = \tau \frac{\bar{N}}{2\hat{n}} + \alpha k \sqrt{R \frac{\bar{N}}{2\hat{n}}} = w(2\hat{n})$$

It follows that $2w(2\hat{n}) > w(\hat{n})$:

$$2w(2\hat{n}) = 2\left(\tau \frac{\bar{N}}{2\hat{n}} + \alpha k \sqrt{R \frac{\bar{N}}{2\hat{n}}}\right) = \tau \frac{\bar{N}}{\hat{n}} + \sqrt{2}\alpha k \sqrt{R \frac{\bar{N}}{\hat{n}}} > \tau \frac{\bar{N}}{\hat{n}} + \alpha k \sqrt{R \frac{\bar{N}}{\hat{n}}} = w(\hat{n})$$

For the number of tours, we show that

$$\widetilde{v}(2\hat{n}-1) = \frac{2\widetilde{w}(2\hat{n}-1)}{l} = \frac{2w(2\hat{n})}{l} > \frac{w(\hat{n})}{l} = v(\hat{n}),$$

For the required vehicle capacity, it follows that

$$\widetilde{Q}(2\hat{n}-1) < Q(\hat{n})$$

Now, we assess the change in expected profit when we move from $\hat{n}$ consecutive to $2\hat{n}-1$ overlapping time windows:

$$\begin{aligned}
\widetilde{P}(2\hat{n}-1) - P(\hat{n}) &= r(2\hat{n}-1)N(2\hat{n}-1) - c2\hat{n}\widetilde{w}(2\hat{n}-1) - \frac{F}{l}2\widetilde{w}(2\hat{n}-1) - P(\hat{n}) \\
&= r\bar{N} - c2\hat{n}w(2\hat{n}) - \frac{F}{l}2w(2\hat{n}) - r\bar{N} + c\hat{n}w(\hat{n}) + \frac{F}{l}w(\hat{n}) \\
&= \left(c\hat{n} + \frac{F}{l}\right)\left(w(\hat{n}) - 2w(2\hat{n})\right) < 0
\end{aligned}$$

□

Proof of Proposition 3.4. Consider Equations (3.1) to (3.4) for a given time window length $l > 0$ and demand $N(n) = \hat{N} \geq N^v(l)$. Let the net margin per order be positive, $r > c\tau$.

1. We first show that the capacity requirements are increasing in n:

$$Q(n) = \frac{n\hat{N}}{v} = nQ(1) \qquad \text{and} \qquad Q(n+1) = (n+1)Q(1) = Q(n) + Q(1)$$

   Now, consider $\hat{n} \geq 1$ consecutive time windows and sufficiently large maximum vehicle capacity $Q(\hat{n}) \leq \bar{Q}$. We first show the necessary and sufficient condition that ensures increasing profits:

$$\begin{aligned} & P(\hat{n}+1) - P(\hat{n}) \overset{!}{>} 0 \\ \Leftrightarrow \quad & r - cw^c(\hat{N}) > 0 \\ \Leftrightarrow \quad & \sqrt{\frac{\hat{N}}{R}} > \frac{c\alpha k}{r - c\tau} \\ \Leftrightarrow \quad & \hat{\delta} > \left(\frac{c\alpha k}{r - c\tau}\right)^2 \quad \Leftrightarrow \quad r - c\tau > \frac{c\alpha k}{\sqrt{\hat{\delta}}} \end{aligned}$$

   with demand density $\hat{\delta} = \frac{\hat{N}}{R}$.

2. Consider $\hat{n} \geq 2$ consecutive time windows, shift length $L(\hat{n}) > \frac{F\tau}{r - c\tau}$, and sufficiently large maximum vehicle capacity $Q(\hat{n}) \leq \bar{Q}$. We first show that the capacity requirements are decreasing:

$$\widetilde{Q}(2\hat{n}-1) = \frac{L(\hat{n})\hat{N}}{\tau\hat{N} + \sqrt{\frac{2\hat{n}}{2\hat{n}-1}}\alpha k\sqrt{R\hat{N}}}$$

   Since $\sqrt{\frac{2\hat{n}}{2\hat{n}-1}} > 1$ it follows that

$$\widetilde{Q}(2\hat{n}-1) = \frac{L(\hat{n})\hat{N}}{\tau\hat{N} + \sqrt{\frac{2\hat{n}}{2\hat{n}-1}}\alpha k\sqrt{R\hat{N}}} < \frac{L(\hat{n})\hat{N}}{\tau\hat{N} + \alpha k\sqrt{R\hat{N}}} = Q(\hat{n})$$

Next, we find the condition for $\hat{N}$ such that $\widetilde{P}(2\hat{n}-1) > P(\hat{n})$. First, we reformulate the profit function for consecutive time windows:

$$\begin{aligned}
P(\hat{n}) &= r\hat{n}N(\hat{n}) - c\hat{n}w(\hat{n}) - \frac{F}{l}w(\hat{n}) \\
&= r\hat{n}N(\hat{n}) - c\hat{n}w(\hat{n}) - \frac{F}{L(\hat{n})}\hat{n}w(\hat{n}) \\
&= \hat{n}r\hat{N} - \hat{n}\left(c + \frac{F}{L(\hat{n})}\right)\left(\tau\hat{N} + \alpha k\sqrt{R\hat{N}}\right) \\
&= \hat{n}r\hat{N} - \frac{C^t(\hat{n})}{L(\hat{n})}\hat{n}\tau\hat{N} - \frac{C^t(\hat{n})}{L(\hat{n})}\hat{n}\alpha k\sqrt{R\hat{N}} \\
&= \frac{rL(\hat{n}) - C^t(\hat{n})\tau}{L(\hat{n})}\hat{n}\hat{N} - \frac{C^t(\hat{n})\alpha k\sqrt{R}}{L(\hat{n})}\hat{n}\sqrt{\hat{N}}
\end{aligned}$$

We then reformulate the profit function for overlapping time windows:

$$\begin{aligned}
\widetilde{P}(2\hat{n}-1) &= r(2\hat{n}-1)N(2\hat{n}-1) - c2\hat{n}\widetilde{w}(2\hat{n}-1) - \frac{F}{l}2\widetilde{w}(2\hat{n}-1) \\
&= r(2\hat{n}-1)\hat{N} - 2\hat{n}c\widetilde{w}(2\hat{n}-1) - \frac{F}{L(\hat{n})}2\hat{n}\widetilde{w}(2\hat{n}-1) \\
&= r(2\hat{n}-1)\hat{N} - 2\hat{n}\left(c + \frac{F}{L(\hat{n})}\right)\left(\tau\frac{2\hat{n}-1}{2\hat{n}}\hat{N} + \alpha k\sqrt{R\frac{2\hat{n}-1}{2\hat{n}}\hat{N}}\right) \\
&= r(2\hat{n}-1)\hat{N} - \frac{C^t(\hat{n})\tau}{L(\hat{n})}(2\hat{n}-1)\hat{N} - \frac{C^t(\hat{n})\alpha k\sqrt{R}}{L(\hat{n})}\sqrt{2\hat{n}(2\hat{n}-1)}\sqrt{\hat{N}} \\
&= \frac{rL(\hat{n}) - C^t(\hat{n})\tau}{L(\hat{n})}(2\hat{n}-1)\hat{N} - \frac{C^t(\hat{n})\alpha k\sqrt{R}}{L(\hat{n})}\sqrt{2\hat{n}(2\hat{n}-1)}\sqrt{\hat{N}}
\end{aligned}$$

The profitability condition becomes

$$\begin{aligned}
&\widetilde{P}(2\hat{n}-1) \overset{!}{>} P(\hat{n}) \\
\Leftrightarrow\quad &(\hat{n}-1)\frac{rL(\hat{n}) - C^t(\hat{n})\tau}{L(\hat{n})}\hat{N} - \left(\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}\right)\frac{C^t(\hat{n})\alpha k\sqrt{R}}{L(\hat{n})}\sqrt{\hat{N}} > 0 \\
\Leftrightarrow\quad &(\hat{n}-1)(rL(\hat{n}) - C^t(\hat{n})\tau)\hat{N} > \left(\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}\right)C^t(\hat{n})\alpha k\sqrt{R\hat{N}} \\
\Leftrightarrow\quad &\sqrt{\frac{\hat{N}}{R}} > \frac{\left(\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}\right)C^t(\hat{n})\alpha k}{(\hat{n}-1)(rL(\hat{n}) - C^t(\hat{n})\tau)}
\end{aligned}$$

We want the right-hand side to be positive. Since $\hat{n} \geq 2$, we only need $rL(\hat{n}) - C^t(\hat{n})\tau$ to be positive. This is achieved for $L(\hat{n}) > \frac{F\tau}{r - c\tau}$ which is a prerequisite of Proposition 3.4.2. Thus, we get

$$\Leftrightarrow \quad \hat{\delta} > \left(\frac{\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}}{\hat{n}-1} \frac{C^t(\hat{n})\alpha k}{rL(\hat{n}) - C^t(\hat{n})\tau} \right)^2 = \left(\frac{\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}}{\hat{n}-1} \right)^2 \delta^P(\hat{n})$$

with demand density $\hat{\delta} = \frac{\hat{N}}{R}$.

We compare this threshold with the threshold derived in Proposition 3.4.1. We see that for $\hat{n} \geq 2$, the term $\frac{\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}}{\hat{n}-1}$ is decreasing in $\hat{n}$ and $\lim_{\hat{n} \to \infty} \frac{\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}}{\hat{n}-1} = 1$. For $\delta^P(\hat{n})$, we get

$$\sqrt{\delta^P(\hat{n})} = \frac{C^t(\hat{n})\alpha k}{rL(\hat{n}) - C^t(\hat{n})\tau} = \frac{(L(\hat{n}) + \frac{F}{c})c\alpha k}{rL(\hat{n}) - (L(\hat{n}) + \frac{F}{c})c\tau} = \frac{c\alpha k}{r\frac{L(\hat{n})}{L(\hat{n}) + \frac{F}{c}} - c\tau}$$

Since $\frac{L(\hat{n})}{L(\hat{n}) + \frac{F}{c}} < 1$, it follows that $r\frac{L(\hat{n})}{L(\hat{n}) + \frac{F}{c}} - c\tau < r - c\tau$. We get

$$\delta^P(\hat{n}) > \left(\frac{c\alpha k}{r - c\tau} \right)^2$$

which concludes

$$\left(\frac{\sqrt{2\hat{n}(2\hat{n}-1)} - \hat{n}}{\hat{n}-1} \right)^2 \delta^P(\hat{n}) > \left(\frac{c\alpha k}{r - c\tau} \right)^2$$

□

Appendix B

Proofs of Chapter IV

Proof of Lemma 4.1. Consider a time window assortment $T \in \widetilde{\mathcal{T}}$ consisting of $n = 3$ overlapping time windows. Let $\mathbf{N} = (N_i \geq 0)_{i=1,\dots,3}$ be a non-negative demand realization with total demand volume $D = \sum_{i=1}^{3} N_i > 0$. We get:

For $n = 3$ the linear program (LP) corresponding to Equation (4.9) is

$$
\begin{aligned}
\min_{(Z,\boldsymbol{\eta})} \quad & Z \\
\text{s.t.} \quad & Z \geq \eta_1 N_1(T), \\
& Z \geq (1-\eta_1) N_1(T) + \eta_2 N_2(T), \\
& Z \geq (1-\eta_2) N_2(T) + \eta_3 N_3(T), \\
& Z \geq (1-\eta_3) N_3(T), \\
& \eta_1, \eta_2, \eta_3 \in [0,1]
\end{aligned}
$$

First, we bring the LP in standard form of

$$
\begin{aligned}
\min_{x} \quad & c^T x \\
\text{s.t.} \quad & Ax \geq b, \\
& x \geq 0
\end{aligned}
$$

We get

$$x = \begin{pmatrix} Z \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}, \quad c = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & -N_1 & 0 & 0 \\ 1 & N_1 & -N_2 & 0 \\ 1 & 0 & N_2 & -N_3 \\ 1 & 0 & 0 & N_3 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 \\ N_1 \\ N_2 \\ N_3 \\ -1 \\ -1 \\ -1 \end{pmatrix}$$

For an LP, the conditions of feasibility, stationarity, and complementary slackness are both necessary and sufficient for optimality. Consider the Lagrange multipliers

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \end{pmatrix} \quad \text{and} \quad \nu = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}$$

We get the optimality conditions

(i) Primal feasibility:

$$Ax \geq b$$
$$x \geq 0$$

(ii) Dual feasibility:

$$\lambda \geq 0$$
$$\nu \geq 0$$

(iii) Stationarity:

$$c - A^T \lambda - \nu = 0$$

(iv) Complementary slackness:

$$\lambda_i (Ax - b)_i = 0 \quad \forall i = 1, \ldots, 7$$
$$\nu_j x_j = 0 \quad \forall j = 1, \ldots, 4$$

resulting in the following system of (in-)equalities

$$Z - N_1\eta_1 \geq 0 \tag{B.1}$$
$$Z + N_1\eta_1 - N_2\eta_2 \geq N_1 \tag{B.2}$$
$$Z + N_2\eta_2 - N_3\eta_3 \geq N_2 \tag{B.3}$$
$$Z + N_3\eta_3 \geq N_3 \tag{B.4}$$
$$-\eta_1 \geq -1 \tag{B.5}$$
$$-\eta_2 \geq -1 \tag{B.6}$$
$$-\eta_3 \geq -1 \tag{B.7}$$
$$x \geq 0 \tag{B.8}$$
$$\lambda \geq 0 \tag{B.9}$$
$$\nu \geq 0 \tag{B.10}$$
$$1 - \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4 - \nu_1 = 0 \tag{B.11}$$
$$N_1\lambda_1 - N_1\lambda_2 + \lambda_5 - \nu_2 = 0 \tag{B.12}$$
$$N_2\lambda_2 - N_2\lambda_3 + \lambda_6 - \nu_3 = 0 \tag{B.13}$$
$$N_3\lambda_3 - N_3\lambda_4 + \lambda_7 - \nu_4 = 0 \tag{B.14}$$
$$\lambda_1(Z - N_1\eta_1) = 0 \tag{B.15}$$
$$\lambda_2(Z + N_1\eta_1 - N_2\eta_2 - N_1) = 0 \tag{B.16}$$
$$\lambda_3(Z + N_2\eta_2 - N_3\eta_3 - N_2) = 0 \tag{B.17}$$
$$\lambda_4(Z + N_3\eta_3 - N_3) = 0 \tag{B.18}$$
$$\lambda_5(1 - \eta_1) = 0 \tag{B.19}$$
$$\lambda_6(1 - \eta_2) = 0 \tag{B.20}$$
$$\lambda_7(1 - \eta_3) = 0 \tag{B.21}$$
$$\nu_1 Z = 0 \tag{B.22}$$
$$\nu_2\eta_1 = 0 \tag{B.23}$$
$$\nu_3\eta_2 = 0 \tag{B.24}$$
$$\nu_4\eta_3 = 0 \tag{B.25}$$

with Equations (1) – (8) primal feasibility, Equations (9) and (10) dual feasibility, Equations (11) – (14) stationarity, and Equations (15) – (25) complementary slackness.

Parameter space: The parameter space in this problem contains the initial demands $N_1, N_2, N_3 \geq 0$ with total demand of $D = \sum_{i=1}^{3} N_i > 0$. Thus, the combination of $\frac{N_1}{D}$ and $\frac{N_3}{D}$ fully defines the parameter space.

Case 1: We first consider the case where the demands of both consecutive time windows, N_1 and N_3, are relatively small, so the overlapping time window holds most of the demand. We show that the solution

$$x = \begin{pmatrix} Z \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}N_2 \\ 1 \\ \frac{1}{2} \\ 0 \end{pmatrix}$$

with $N_2 > 0$ satisfies the optimality conditions and we provide the parameter bounds for the solution to be feasible.

From the variable values, we get $\nu_1 = \nu_2 = \nu_3 = 0$ and the inactive constraints result in $\lambda_1 = \lambda_4 = \lambda_6 = \lambda_7 = 0$. Stationarity equations (11) and (13) require $\lambda_2 = \lambda_3 = \frac{1}{2}$. Stationarity equation (12) requires $\lambda_5 = \frac{1}{2}N_1$, and equation (14) implies $\nu_4 = \frac{1}{2}N_3$. Thus, the solution fulfills stationarity, complementary slackness, and is dual feasible.

We look at the remaining constraints of primal feasibility to determine the feasible region within the parameter space. Constraint (1) leads to the condition on N_1

$$0 \leq \frac{N_1}{D} \leq \frac{1}{3}\left(1 - \frac{N_3}{D}\right)$$

and constraint (3) provides an upper bound on N_3

$$0 \leq \frac{N_3}{D} \leq \frac{1}{3}\left(1 - \frac{N_1}{D}\right).$$

The conditions together ensure $N_2 > 0$.

<u>Case 2.1:</u> In this case, N_1 and N_2 share most of the demand, leaving N_3 relatively small. We show that the solution

$$x = \begin{pmatrix} Z \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3}(N_2 + N_1) \\ \frac{1}{3}\left(1 + \frac{N_2}{N_1}\right) \\ \frac{1}{3}\left(2 - \frac{N_1}{N_2}\right) \\ 0 \end{pmatrix}$$

with $N_1, N_2 > 0$ satisfies the optimality conditions and we provide the parameter bounds for the solution to be feasible.

From the variable values, we get $\nu_1 = \nu_2 = 0$ and the inactive constraints result in $\lambda_4 = \lambda_6 = \lambda_7 = 0$. Further, a choice of $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3}$, $\lambda_5 = 0$, $\nu_3 = 0$, and $\nu_4 = \frac{1}{3}N_3$ satisfies the stationarity equations (11) – (14). Thus, the solution fulfills stationarity, complementary slackness, and is dual feasible.

We look at the remaining constraints of primal feasibility to determine the feasible region within the parameter space. Constraint (4) leads to the condition on N_3

$$0 \leq \frac{N_3}{D} \leq \frac{1}{4}$$

and constraint (5) and non-negativity of η_2 in constraint (8) provide the bounds on N_1

$$\frac{1}{3}\left(1 - \frac{N_3}{D}\right) \leq \frac{N_1}{D} \leq \frac{2}{3}\left(1 - \frac{N_3}{D}\right).$$

The conditions together ensure $N_1, N_2 > 0$.

<u>Case 2.2:</u> This case follows from the symmetry of the problem. We get the solution

$$x = \begin{pmatrix} Z \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3}(N_2 + N_3) \\ 1 \\ \frac{1}{3}\left(1 + \frac{N_3}{N_2}\right) \\ \frac{1}{3}\left(2 - \frac{N_2}{N_3}\right) \end{pmatrix}$$

with bounds that ensure $N_2, N_3 > 0$:

$$0 \leq \frac{N_1}{D} \leq \frac{1}{4} \quad \text{and} \quad \frac{1}{3}\left(1 - \frac{N_1}{D}\right) \leq \frac{N_3}{D} \leq \frac{2}{3}\left(1 - \frac{N_1}{D}\right).$$

Case 3: In this case, all demands are sufficiently balanced such that the optimal allocation results in homogeneous demand. We show that the solution

$$x = \begin{pmatrix} Z \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{4}D \\ \frac{1}{4}\left(1 + \frac{N_2}{N_1} + \frac{N_3}{N_1}\right) \\ \frac{1}{2}\left(1 - \frac{N_1}{N_2} + \frac{N_3}{N_2}\right) \\ \frac{1}{4}\left(3 - \frac{N_1}{N_3} - \frac{N_2}{N_3}\right) \end{pmatrix}$$

with $N_1, N_2, N_3 > 0$ satisfies the optimality conditions and we provide the parameter bounds for the solution to be feasible.

From the variable values, we get $\nu_1 = \nu_2 = 0$ and the inactive constraints result in $\lambda_7 = 0$. Further, a choice of $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{4}$, $\lambda_5 = \lambda_6 = 0$, $\nu_3 = \nu_4 = 0$ satisfies the stationarity equations (11) – (14). Thus, the solution fulfills stationarity, complementary slackness, and is dual feasible.

We look at the remaining constraints of primal feasibility to determine the feasible region within the parameter space. Constraint (5) and non-negativity of η_2 in constraint (8) leads to the bounds on N_1

$$\frac{1}{4} \leq \frac{N_1}{D} \leq \frac{1}{2}.$$

Constraint (6) and non-negativity of η_3 in constraint (8) provide the bounds on N_3

$$\frac{1}{4} \leq \frac{N_3}{D} \leq \frac{1}{2}.$$

The conditions together ensure $N_1, N_2, N_3 > 0$.

Cases 4.1 and 4.2: There are only two possible symmetric cases left, where either N_1 or N_3, respectively, make at least 50% of the total demand volume alone. N_2 is small enough such that cases 2.1 and 2.2 do not apply. The maximum time window demands take the values $\frac{1}{2}N_1$ and $\frac{1}{2}N_3$, respectively and the parameter bounds cover the remaining parts of the parameter space.

□

Proof of Lemma 4.2. Consider Equations (4.2) and (4.7), let the corresponding demand vectors $\boldsymbol{\mathcal{N}}^T$ follow a distribution with fixed total demand volume D, and define $w^D := \tau D + \alpha k \sqrt{RD}$. We establish bounds on the relevant bottleneck workloads for consecutive and overlapping time windows, respectively.

a) Consider an assortment $T \in \mathcal{T}$ consisting of n consecutive time windows. It follows directly from Corollary 4.1 that the best-case demand realization $\mathbf{N}^{T+}$ fulfills $N_i^{T+} = \frac{D}{n}, \forall i = 1, \dots n$. Then $\max_{i=1,\dots,n} w_i(\mathbf{N}^{T+}) = w\left(\frac{D}{n}\right)$. The worst-case demand realization $\mathbf{N}^{T-}$ fulfills $N_k^{T-} = D$ for a $k \in \{1, \dots, n\}$. Then $\max_{i=1,\dots,n} w_i(\mathbf{N}^{T-}) = w(D)$.

   The upper bound results from the worst-case demand realization and is

$$w(D) =: w^D$$

   and the lower bound results from the best-case demand realization and is

$$\begin{aligned} w\left(\frac{D}{n}\right) &= \tau \frac{D}{n} + \alpha k \sqrt{R \frac{D}{n}} \\ &= \frac{1}{\sqrt{n}} \left[\tau D + \alpha k \sqrt{RD}\right] - \left(\frac{1}{\sqrt{n}} - \frac{1}{n}\right) \tau D \\ &= \frac{1}{\sqrt{n}} w^D - \left(\frac{1}{\sqrt{n}} - \frac{1}{n}\right) \tau D \end{aligned}$$

b) Consider an assortment $T \in \tilde{\mathcal{T}}$ consisting of n overlapping time windows. A best-case demand realization $\mathbf{N}^{T+}$ needs to enable that the allocated demand is evenly distributed, i.e., $\theta_i(\boldsymbol{\eta}^*, \mathbf{N}^{T+}) = \frac{D}{n+1}, \forall i = 1, \dots, n+1$ under optimal allocation weights $\boldsymbol{\eta}^*$. This follows directly from Corollary 4.1. Then, $2 \cdot R(T, \mathbf{N}^{T+}) = 2 \cdot w\left(\frac{nN^*}{n+1}\right)$. For a worst-case demand realization $\mathbf{N}^{T-}$, according to Equation (4.4), the total demand can only be distributed over two delivery intervals, resulting in $\theta_k(\boldsymbol{\eta}^*, \mathbf{N}^{T-}) = \frac{D}{2}$ for some $k = 1, \dots, n+1$. Then $2 \cdot R(T, \mathbf{N}^{T-}) = 2 \cdot w\left(\frac{D}{2}\right)$.

Consequently, the upper bound, resulting from the worst-case demand realization, becomes

$$\begin{aligned} 2 \cdot w\left(\frac{D}{2}\right) &= \tau D + \alpha k \sqrt{2RD} \\ &= \sqrt{2}\left[\tau D + \alpha k \sqrt{RD}\right] - \left(\sqrt{2}-1\right)\tau D \\ &= \sqrt{2} w^D - \left(\sqrt{2}-1\right)\tau D \end{aligned}$$

and the lower bound, resulting form the best-case demand realization, becomes

$$\begin{aligned} 2 \cdot w\left(\frac{D}{n+1}\right) &= 2\tau \frac{D}{n+1} + 2\alpha k \sqrt{R\frac{D}{n+1}} \\ &= \frac{2}{\sqrt{n+1}}\left[\tau D + \alpha k \sqrt{RD}\right] - \left(\frac{2}{\sqrt{n+1}} - \frac{2}{n+1}\right)\tau D \\ &= \frac{2}{\sqrt{n+1}} w^D - \left(\frac{2}{\sqrt{n+1}} - \frac{2}{n+1}\right)\tau D \end{aligned}$$

□

Proof of Theorem 4.1. Consider Equations (4.3) and (4.8) for any assortments $T_{\hat{n}} \in \mathcal{T}$ and $T_{2\hat{n}-1} \in \widetilde{\mathcal{T}}$, each comprising $\hat{n}$ consecutive time windows of length l. Let $\mathbf{N}^{T_{\hat{n}}}$ and $\mathbf{N}^{T_{2\hat{n}-1}}$ denote any respective demand realizations. Define the bottleneck demand $\bar{N} := \max_{i=1,\dots,\hat{n}} N_i^{T_{\hat{n}}}$ and the allocated bottleneck demand $\bar{\theta} := \max_{i=1,\dots,2\hat{n}} \theta_i(\boldsymbol{\eta}^*, \mathbf{N}^{T_{2\hat{n}-1}})$ under the optimal allocation $\boldsymbol{\eta}^*$.

The allocated bottleneck demand, which can be expressed by $\bar{\theta} = \phi\bar{N}$, for which the overlapping assortment yields lower total delivery costs than the consecutive assortment, satisfies

$$\begin{aligned} & 2w(\phi\bar{N}) < w(\bar{N}) \\ \Leftrightarrow\quad & 2\tau\phi\bar{N} + 2\alpha k\sqrt{R\phi\bar{N}} < \tau\bar{N} + \alpha k\sqrt{R\bar{N}} \\ \Leftrightarrow\quad & \tau\left(2\phi\bar{N} - \bar{N}\right) + \alpha k\sqrt{R}\left(\sqrt{4\phi\bar{N}} - \sqrt{\bar{N}}\right) < 0. \end{aligned}$$

Solving for ϕ, we get

$$\phi < \frac{\tau\alpha k\sqrt{R\bar{N}} + \tau^2\bar{N} + (\alpha k)^2 R - \alpha k\sqrt{R}\sqrt{2\tau\alpha k\sqrt{R\bar{N}} + 2\tau^2\bar{N} + (\alpha k)^2 R}}{2\tau^2\bar{N}}$$

$$\Leftrightarrow \quad \phi < \frac{\tau\left(\tau + \alpha k\sqrt{\frac{R}{\bar{N}}}\right) + \left(\alpha k\sqrt{\frac{R}{\bar{N}}}\right)^2 - \alpha k\sqrt{\frac{R}{\bar{N}}}\sqrt{2\tau\left(\tau + \alpha k\sqrt{\frac{R}{\bar{N}}}\right) + \left(\alpha k\sqrt{\frac{R}{\bar{N}}}\right)^2}}{2\tau^2}.$$

Defining $d(\bar{N}) = \alpha k\sqrt{\frac{R}{\bar{N}}}$, we get

$$\begin{aligned}
&\phi < \frac{\tau\left(\tau + d(\bar{N})\right) + d(\bar{N})^2 - d(\bar{N})\sqrt{2\tau\left(\tau + d(\bar{N})\right) + d(\bar{N})^2}}{2\tau^2} \\
\Leftrightarrow \quad &\phi < \frac{\tau^2 + \tau d(\bar{N}) + d(\bar{N})^2 - d(\bar{N})\sqrt{2\tau\left(\tau + d(\bar{N})\right) + d(\bar{N})^2}}{2\tau^2} \\
\Leftrightarrow \quad &\phi < \frac{1}{2} + \frac{d(\bar{N})}{2\tau^2}\left((\tau + d(\bar{N})) - \sqrt{2\tau\left(\tau + d(\bar{N})\right) + d(\bar{N})^2}\right) \\
\Leftrightarrow \quad &\phi < \frac{1}{2} - \frac{d(\bar{N})}{2\tau^2}\left(\sqrt{2\tau^2 + 2\tau d(\bar{N}) + d(\bar{N})^2} - (\tau + d(\bar{N}))\right) \\
\Leftrightarrow \quad &\phi < \frac{1}{2} - \frac{d(\bar{N})}{2\tau^2}\left(\sqrt{(\tau + d(\bar{N}))^2 + \tau^2)} - (\tau + d(\bar{N}))\right).
\end{aligned}$$

Defining $w^c(\bar{N}) = \tau + d(\bar{N})$, we get

$$\begin{aligned}
&\phi < \frac{1}{2} - \frac{d(\bar{N})}{2\tau^2}\left(\sqrt{w^c(\bar{N})^2 + \tau^2)} - w^c(\bar{N})\right) \\
\Leftrightarrow \quad &\frac{\bar{\theta}}{\bar{N}} < \frac{1}{2} - \frac{d(\bar{N})}{2\tau^2}\left(\sqrt{w^c(\bar{N})^2 + \tau^2)} - w^c(\bar{N})\right)
\end{aligned}$$

□

Proof of Lemma 4.3. Let $T \in \tilde{\mathcal{T}}$ be any overlapping assortment, $\mathbf{N}^T$ a corresponding demand realization, and consider the allocation mechanism defined in Equation (4.4). Let $\hat{N} := \max_{i=1,\ldots,n} N_i^T$ denote the bottleneck demand, and let $\bar{\theta} := \max_{i=1,\ldots,n+1} \theta_i(\boldsymbol{\eta}^*, \mathbf{N}^T)$ be the allocated bottleneck demand under optimal allocation $\boldsymbol{\eta}^*$.

We aim to show that under any feasible allocation vector $\boldsymbol{\eta} \in [0,1]^n$, the maximum allocated demand satisfies

$$\bar{\theta} := \max_{i=1,\ldots,n+1} \theta_i(\boldsymbol{\eta}, \mathbf{N}^T) \geq \frac{1}{2}\hat{N}, \quad \text{where } \hat{N} := \max_{i=1,\ldots,n} N_i^T.$$

Each demand N_i^T contributes to two adjacent intervals:

- $\eta_i N_i^T$ is allocated to θ_i,

- $(1-\eta_i)N_i^T$ is allocated to θ_{i+1}.

Thus, the total contribution of N_i^T to allocations is

$$\theta_i + \theta_{i+1} \geq \eta_i N_i^T + (1-\eta_i)N_i^T = N_i^T.$$

Assume for contradiction that all allocations are strictly below half the bottleneck demand:

$$\theta_j < \frac{1}{2}\hat{N} \quad \text{for all } j = 1, \ldots, n+1.$$

Let $k \in \{1, \ldots, n\}$ be such that $N_k^T = \hat{N}$. Then:

$$\theta_k + \theta_{k+1} \geq N_k^T = \hat{N},$$

but also, by assumption,

$$\theta_k + \theta_{k+1} < \frac{1}{2}\hat{N} + \frac{1}{2}\hat{N} = \hat{N},$$

a contradiction. Therefore, there exists at least one j such that

$$\theta_j \geq \frac{1}{2}\hat{N}.$$

Hence,

$$\bar{\theta} = \max_j \theta_j \geq \frac{1}{2}\hat{N}.$$

□

Appendix C

Extended Numerical Results of Chapter IV

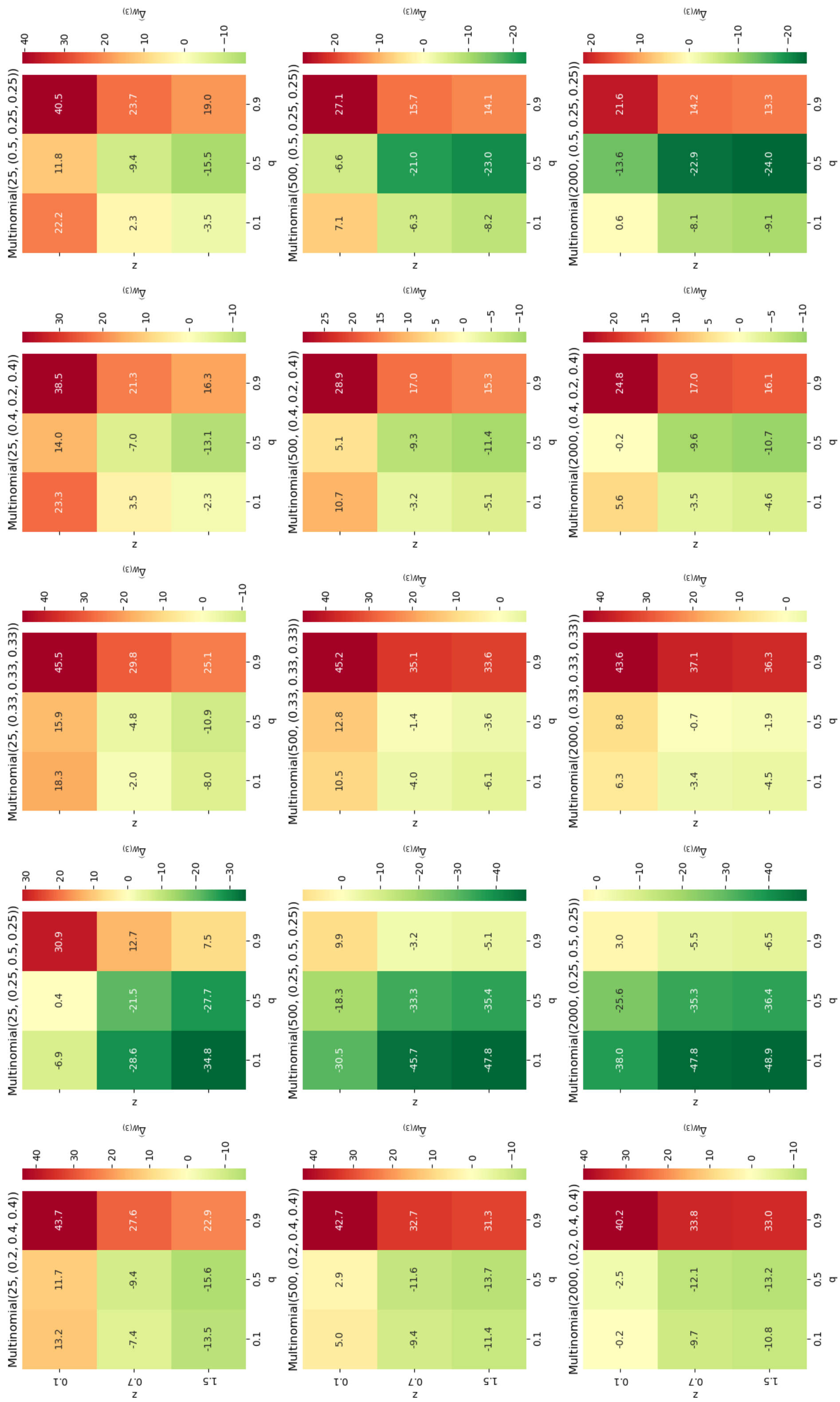


Figure C.1.: Monte Carlo simulation results for $\hat{n} = 3$ consecutive time windows

Bibliography

Agatz, N., Campbell, A., Fleischmann, M., & Savelsbergh, M. (2011). Time Slot Management in Attended Home Delivery. *Transportation Science*, *45*(3), 435–449. https://doi.org/10.1287/trsc.1100.0346

Agatz, N., Campbell, A. M., Fleischmann, M., Van Nunen, J., & Savelsbergh, M. (2013). Revenue management opportunities for Internet retailers. *Journal of Revenue and Pricing Management*, *12*(2), 128–138. https://doi.org/10.1057/rpm.2012.51

Agatz, N., Fan, Y., & Stam, D. (2021). The Impact of Green Labels on Time Slot Choice and Operational Sustainability. *Production and Operations Management*, *30*(7), 2285–2303. https://doi.org/10.1111/poms.13368

Agatz, N. A. H., Fleischmann, M., & Van Nunen, J. A. E. E. (2008). E-fulfillment and multi-channel distribution – A review. *European Journal of Operational Research*, *187*(2), 339–356. https://doi.org/10.1016/j.ejor.2007.04.024

Amorim, P., DeHoratius, N., Eng-Larsson, F., & Martins, S. (2024). Customer Preferences for Delivery Service Attributes in Attended Home Delivery. *Management Science*, *70*(11), 7559–7578. https://doi.org/10.1287/mnsc.2020.01274

Ansari, S., Başdere, M., Li, X., Ouyang, Y., & Smilowitz, K. (2018). Advancements in continuous approximation models for logistics and transportation systems: 1996–2016. *Transportation Research Part B: Methodological*, *107*, 229–252. https://doi.org/10.1016/j.trb.2017.09.019

Asdemir, K., Jacob, V. S., & Krishnan, R. (2009). Dynamic pricing of multiple home delivery options. *European Journal of Operational Research*, *196*(1), 246–257. https://doi.org/10.1016/j.ejor.2008.03.005

Avraham, E., & Raviv, T. (2021). The steady-state mobile personnel booking problem. *Transportation Research Part B: Methodological*, *154*, 266–288. https://doi.org/10.1016/j.trb.2021.10.008

Azi, N., Gendreau, M., & Potvin, J.-Y. (2012). A dynamic vehicle routing problem with multiple delivery routes. *Annals of Operations Research*, *199*(1), 103–112. https://doi.org/10.1007/s10479-011-0991-3

Banerjee, D., Erera, A. L., & Toriello, A. (2022). Fleet Sizing and Service Region Partitioning for Same-Day Delivery Systems. *Transportation Science*, *56*(5), 1327–1347. https://doi.org/10.1287/trsc.2022.1125

Beardwood, J., Halton, J. H., & Hammersley, J. M. (1959). The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, *55*(4), 299–327. https://doi.org/10.1017/S0305004100034095

Belanche, D., Casaló, L. V., Flavián, C., & Pérez-Rueda, A. (2021). The role of customers in the gig economy: How perceptions of working conditions and service quality influence the use and recommendation of food delivery services. *Service Business*, *15*(1), 45–75. https://doi.org/10.1007/s11628-020-00432-7

Belavina, E., Girotra, K., & Kabra, A. (2017). Online Grocery Retail: Revenue Models and Environmental Impact. *Management Science*, *63*(6), 1781–1799. https://doi.org/10.1287/mnsc.2016.2430

Belobaba, P. (1987). *Air travel demand and airline* (FTL Report No. R87-7). Flight Transportation Laboratory, Massachusetts Institute of Technology. Cambridge, MA.

Bernstein, F., Modaresi, S., & Sauré, D. (2019). A Dynamic Clustering Approach to Data-Driven Assortment Personalization. *Management Science*, *65*(5), 2095–2115. https://doi.org/10.1287/mnsc.2018.3031

Birolini, S., Jacquillat, A., Cattaneo, M., & Antunes, A. P. (2021). Airline Network Planning: Mixed-integer non-convex optimization with demand–supply interactions. *Transportation Research Part B: Methodological*, *154*, 100–124. https://doi.org/10.1016/j.trb.2021.09.003

Bjørgen, A., Bjerkan, K. Y., & Hjelkrem, O. A. (2021). E-groceries: Sustainable last mile distribution in city planning. *Research in Transportation Economics*, *87*, 100805. https://doi.org/10.1016/j.retrec.2019.100805

Boyer, K. K., Prud'homme, A. M., & Chung, W. (2009). The Last Mile Challenge: Evaluating The Effects of Customer Density and Delivery Window Patterns. *Journal of Business Logistics*, *30*(1), 185–201. https://doi.org/10.1002/j.2158-1592.2009.tb00104.x

Bruck, B. P., Castegini, F., Cordeau, J.-F., Iori, M., Poncemi, T., & Vezzali, D. (2020). A Decision Support System for Attended Home Services. *INFORMS Journal on Applied Analytics*, *50*(2), 137–152. https://doi.org/10.1287/inte.2020.1031

Bruck, B. P., Cordeau, J.-F., & Iori, M. (2018). A practical time slot management and routing problem for attended home services. *Omega*, *81*, 208–219. https://doi.org/10.1016/j.omega.2017.11.003

Cadarso, L., Vaze, V., Barnhart, C., & Marín, Á. (2017). Integrated Airline Scheduling: Considering Competition Effects and the Entry of the High Speed Rail. *Transportation Science*, *51*(1), 132–154. https://doi.org/10.1287/trsc.2015.0617

Campbell, A. M., & Savelsbergh, M. (2006). Incentive Schemes for Attended Home Delivery Services. *Transportation Science*, *40*(3), 327–341. https://doi.org/10.1287/trsc.1050.0136

Campbell, A. M., & Savelsbergh, M. W. P. (2005). Decision Support for Consumer Direct Grocery Initiatives. *Transportation Science*, *39*(3), 313–327. https://doi.org/10.1287/trsc.1040.0105

Casazza, M., Ceselli, A., & Létocart, L. (2016). Optimizing Time Slot Allocation in Single Operator Home Delivery Problems. In M. Lübbecke, A. Koster, P. Letmathe, R. Madlener, B. Peis, & G. Walther (Eds.), *Operations Research Proceedings 2014* (pp. 91–97). Springer International Publishing. https://doi.org/10.1007/978-3-319-28697-6_14

Çeven, E., & Gue, K. R. (2017). Optimal Wave Release Times for Order Fulfillment Systems with Deadlines. *Transportation Science*, *51*(1), 52–66. https://doi.org/10.1287/trsc.2015.0642

Chen, X., Wang, T., Thomas, B. W., & Ulmer, M. W. (2023). Same-day delivery with fair customer service. *European Journal of Operational Research*, *308*(2), 738–751. https://doi.org/10.1016/j.ejor.2022.12.009

Chung, H., Ahn, H.-S., & Jasin, S. (2019). (Rescaled) Multi-Attempt Approximation of Choice Model and Its Application to Assortment Optimization. *Production and Operations Management*, *28*(2), 341–353. https://doi.org/10.1111/poms.12916

Cleophas, C., & Ehmke, J. F. (2014). When Are Deliveries Profitable?: Considering Order Value and Transport Capacity in Demand Fulfillment for Last-Mile Deliveries in Metropolitan Areas. *Business & Information Systems Engineering*, *6*(3), 153–163. https://doi.org/10.1007/s12599-014-0321-9

Coltman, T., & Devinney, T. M. (2013). Modeling the operational capabilities for customized and commoditized services. *Journal of Operations Management*, *31*(7-8), 555–566. https://doi.org/10.1016/j.jom.2013.09.002

Cordeau, J.-F., Iori, M., & Vezzali, D. (2023). A survey of attended home delivery and service problems with a focus on applications. *4OR*, *21*(4), 547–583. https://doi.org/10.1007/s10288-023-00556-2

Côté, J.-F., Mansini, R., & Raffaele, A. (2019). Tactical Time Window Management in Attended Home Delivery. *Working Paper*, *FSA-2019-005*.

Daganzo, C. (2005). *Logistics systems analysis: 4 tables* (4. ed). Springer.

Daganzo, C. F. (1984). The Distance Traveled to Visit N Points with a Maximum of C Stops per Vehicle: An Analytic Model and an Application. *Transportation Science*, *18*(4), 331–350. https://doi.org/10.1287/trsc.18.4.331

Daganzo, C. F. (1987). Modeling Distribution Problems with Time Windows: Part I. *Transportation Science*, *21*(3), 171–179. https://doi.org/10.1287/trsc.21.3.171

Dayarian, I., & Pazour, J. (2022). Crowdsourced order-fulfillment policies using in-store customers. *Production and Operations Management*, *31*(11), 4075–4094. https://doi.org/10.1111/poms.13805

Delasay, M., Jain, A., & Kumar, S. (2022). Impacts of the COVID-19 pandemic on grocery retail operations: An analytical model. *Production and Operations Management*, *31*(5), 2237–2255. https://doi.org/10.1111/poms.13717

Ehmke, J. F., & Campbell, A. M. (2014). Customer acceptance mechanisms for home deliveries in metropolitan areas. *European Journal of Operational Research*, *233*(1), 193–207. https://doi.org/10.1016/j.ejor.2013.08.028

Feldman, J., & Paul, A. (2019). Relating the Approximability of the Fixed Cost and Space Constrained Assortment Problems. *Production and Operations Management*, *28*(5), 1238–1255. https://doi.org/10.1111/poms.12983

Feldman, J., & Topaloglu, H. (2015). Bounding Optimal Expected Revenues for Assortment Optimization under Mixtures of Multinomial Logits. *Production and Operations Management*, *24*(10), 1598–1620. https://doi.org/10.1111/poms.12365

Feng, Q., Li, C., Lu, M., & Shanthikumar, J. G. (2022). Dynamic Substitution for Selling Multiple Products under Supply and Demand Uncertainties. *Production and Operations Management*, *31*(4), 1645–1662. https://doi.org/10.1111/poms.13636

Feng, Q., Shanthikumar, J. G., & Xue, M. (2022). Consumer Choice Models and Estimation: A Review and Extension. *Production and Operations Management*, *31*(2), 847–867. https://doi.org/10.1111/poms.13499

Fikar, C., Mild, A., & Waitz, M. (2021). Facilitating consumer preferences and product shelf life data in the design of e-grocery deliveries. *European Journal of Operational Research*, *294*(3), 976–986. https://doi.org/10.1016/j.ejor.2019.09.039

Fleckenstein, D., Klein, R., Klein, V., & Steinhardt, C. (2025). On the Concept of Opportunity Cost in Integrated Demand Management and Vehicle Routing. *Transportation Science*, *59*(1), 125–142. https://doi.org/10.1287/trsc.2024.0644

Fleckenstein, D., Klein, R., & Steinhardt, C. (2023). Recent advances in integrating demand management and vehicle routing: A methodological review. *European Journal of Operational Research*, *306*(2), 499–518. https://doi.org/10.1016/j.ejor.2022.04.032

Fleischmann, B., Meyr, H., & Wagner, M. (2015). Advanced Planning. In H. Stadtler, C. Kilger, & H. Meyr (Eds.), *Supply Chain Management and Advanced Planning* (pp. 71–95). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-55309-7_4

Goethals, F., Leclercq-Vandelannoitte, A., & Tütüncü, Y. (2012). French consumers' perceptions of the unattended delivery model for e-grocery retailing. *Journal of Retailing and Consumer Services*, *19*(1), 133–139. https://doi.org/10.1016/j.jretconser.2011.11.002

Gümüş, M., Li, S., Oh, W., & Ray, S. (2013). Shipping Fees or Shipping Free? A Tale of Two Price Partitioning Strategies in Online Retailing. *Production and Operations Management*, *22*(4), 758–776. https://doi.org/10.1111/j.1937-5956.2012.01391.x

Han, S., Zhao, L., Chen, K., Luo, Z.-w., & Mishra, D. (2017). Appointment scheduling and routing optimization of attended home delivery system with random customer behavior. *European Journal of Operational Research*, *262*(3), 966–980. https://doi.org/10.1016/j.ejor.2017.03.060

Haugland, D., Ho, S. C., & Laporte, G. (2007). Designing delivery districts for the vehicle routing problem with stochastic demands. *European Journal of Operational Research*, *180*(3), 997–1010. https://doi.org/10.1016/j.ejor.2005.11.070

Heger, J., & Klein, R. (2024). Assortment optimization: A systematic literature review. *OR Spectrum*, *46*, 1099–1161. https://doi.org/10.1007/s00291-024-00752-4

Hernandez, F., Gendreau, M., & Potvin, J.-Y. (2017). Heuristics for tactical time slot management: A periodic vehicle routing problem view. *International Transactions in Operational Research*, *24*(6), 1233–1252. https://doi.org/10.1111/itor.12403

Hübner, A., Schäfer, F., & Schaal, K. N. (2020). Maximizing Profit via Assortment and Shelf-Space Optimization for Two-Dimensional Shelves. *Production and Operations Management*, *29*(3), 547–570. https://doi.org/10.1111/poms.13111

Hungerländer, P., Rendl, A., & Truden, C. (2017). On the Slot Optimization Problem in On-Line Vehicle Routing. *Transportation Research Procedia*, *27*, 492–499. https://doi.org/10.1016/j.trpro.2017.12.046

Hungerländer, P., & Truden, C. (2018). Efficient and Easy-to-Implement Mixed-Integer Linear Programs for the Traveling Salesperson Problem with Time Windows. *Transportation Research Procedia*, *30*, 157–166. https://doi.org/10.1016/j.trpro.2018.09.018

Jiang, X., Wang, H., Guo, X., & Gong, X. (2019). Using the FAHP, ISM, and MICMAC Approaches to Study the Sustainability Influencing Factors of the Last Mile Delivery of Rural E-Commerce Logistics. *Sustainability*, *11*(14), 3937. https://doi.org/10.3390/su11143937

Jimenez G, H. S., Rodrigues, T. F., Dantas, M. M., & Cavalcante, C. A. (2020). A dynamic inventory rationing policy for business-to-consumer e-tail stores in a supply disruption context. *Computers & Industrial Engineering*, *142*, 106379. https://doi.org/10.1016/j.cie.2020.106379

Klapp, M. A., Erera, A. L., & Toriello, A. (2018). The Dynamic Dispatch Waves Problem for same-day delivery. *European Journal of Operational Research*, *271*(2), 519–534. https://doi.org/10.1016/j.ejor.2018.05.032

Klapp, M. A., Erera, A. L., & Toriello, A. (2020). Request acceptance in same-day delivery. *Transportation Research Part E: Logistics and Transportation Review*, *143*, 102083. https://doi.org/10.1016/j.tre.2020.102083

Klein, R., Koch, S., Steinhardt, C., & Strauss, A. K. (2020). A review of revenue management: Recent generalizations and advances in industry applications. *European Journal of Operational Research*, *284*(2), 397–412. https://doi.org/10.1016/j.ejor.2019.06.034

Klein, R., Mackert, J., Neugebauer, M., & Steinhardt, C. (2018). A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum*, *40*(4), 969–996. https://doi.org/10.1007/s00291-017-0501-3

Klein, R., Neugebauer, M., Ratkovitch, D., & Steinhardt, C. (2019). Differentiated Time Slot Pricing Under Routing Considerations in Attended Home Delivery. *Transportation Science*, *53*(1), 236–255. https://doi.org/10.1287/trsc.2017.0738

Klein, V., & Steinhardt, C. (2023). Dynamic demand management and online tour planning for same-day delivery. *European Journal of Operational Research*, *307*(2), 860–886. https://doi.org/10.1016/j.ejor.2022.09.011

Koch, S., & Klein, R. (2020). Route-based approximate dynamic programming for dynamic pricing in attended home delivery. *European Journal of Operational Research*, *287*(2), 633–652. https://doi.org/10.1016/j.ejor.2020.04.002

Köhler, C., Ehmke, J. F., & Campbell, A. M. (2020). Flexible time window management for attended home deliveries. *Omega*, *91*, 102023. https://doi.org/10.1016/j.omega.2019.01.001

Köhler, C., & Haferkamp, J. (2019). Evaluation of delivery cost approximation for attended home deliveries. *Transportation Research Procedia*, *37*, 67–74. https://doi.org/10.1016/j.trpro.2018.12.167

Kök, A. G., Fisher, M. L., & Vaidyanathan, R. (2015). Assortment Planning: Review of Literature and Industry Practice. In N. Agrawal & S. A. Smith (Eds.), *Retail Supply Chain Management* (pp. 175–236, Vol. 223). Springer US. https://doi.org/10.1007/978-1-4899-7562-1_8

Lang, M. A. K., & Cleophas, C. (2020). Establishing an Extendable Benchmarking Framework for E-Fulfillment. *Proceedings of the 53rd Hawaii International Conference on System Sciences.*

Lang, M. A. K., Cleophas, C., & Ehmke, J. F. (2021). Anticipative Dynamic Slotting for Attended Home Deliveries. *Operations Research Forum*, *2*, 70. https://doi.org/10.1007/s43069-021-00086-9

Lang, M. A., Cleophas, C., & Ehmke, J. F. (2021). Multi-criteria decision making in dynamic slotting for attended home deliveries. *Omega*, *102*, 102305. https://doi.org/10.1016/j.omega.2020.102305

Le, C., & Mai, T. (2024). Constrained Assortment Optimization under the Cross-Nested Logit Model. *Production and Operations Management*, *33*(10), 2073–2090. https://doi.org/10.1177/10591478241263857

Lebedev, D., Goulart, P., & Margellos, K. (2021). A dynamic programming framework for optimal delivery time slot pricing. *European Journal of Operational Research*, *292*(2), 456–468. https://doi.org/10.1016/j.ejor.2020.11.010

Li, F., Fan, Z.-P., Cao, B.-B., & Li, X. (2021). Logistics Service Mode Selection for Last Mile Delivery: An Analysis Method Considering Customer Utility and Delivery Service Cost. *Sustainability*, *13*, 284. https://doi.org/10.3390/su13010284

Li, G., Sheng, L., & Zhan, D. (2023). Designing shipping policies with top-up options to qualify for free delivery. *Production and Operations Management*, *32*(9), 2704–2722. https://doi.org/10.1111/poms.14002

Lin, I. I., & Mahmassani, H. S. (2002). Can Online Grocers Deliver? Some Logistics Considerations. *Transportation Research Record: Journal of the Transportation Research Board*, *1817*(1), 17–24. https://doi.org/10.3141/1817-03

Liu, N., Van De Ven, P. M., & Zhang, B. (2019). Managing Appointment Booking Under Customer Choices. *Management Science*, *65*(9), 4280–4298. https://doi.org/10.1287/mnsc.2018.3150

Mackert, J. (2019a). Integrating Customer Choice in Differentiated Slotting for Last-Mile Logistics. *Logistics Research*, *12*(1), 1–23. https://doi.org/10.23773/2019_5

Mackert, J. (2019b). Choice-based dynamic time slot management in attended home delivery. *Computers & Industrial Engineering*, *129*, 333–345. https://doi.org/10.1016/j.cie.2019.01.048

Magalhães, D. J. A. V. D. (2021). Analysis of critical factors affecting the final decision-making for online grocery shopping. *Research in Transportation Economics*, *87*, 101088. https://doi.org/10.1016/j.retrec.2021.101088

Manerba, D., Mansini, R., & Zanotti, R. (2018). Attended Home Delivery: Reducing last-mile environmental impact by changing customer habits. *IFAC-PapersOnLine*, *51*(5), 55–60. https://doi.org/10.1016/j.ifacol.2018.06.199

McKinsey. (2022, April). The next horizon for grocery e-commerce [Available at https://www.mckinsey.com/industries/retail/our-insights/the-next-horizon-for-grocery-ecommerce-beyond-the-pandemic-bump (accessed July 18, 2024)].

Metters, R., & Walton, S. (2007). Strategic supply chain choices for multi-channel Internet retailers. *Service Business*, *1*(4), 317–331. https://doi.org/10.1007/s11628-006-0016-5

Meyr, H. (2009). Customer segmentation, allocation planning and order promising in make-to-stock production. *OR Spectrum*, *31*(1), 229–256. https://doi.org/10.1007/s00291-008-0123-x

Milioti, C., Pramatari, K., & Zampou, E. (2020). Choice of prevailing delivery methods in e-grocery: A stated preference ranking experiment. *International Journal of Retail & Distribution Management*, *49*(2), 281–298. https://doi.org/10.1108/IJRDM-08-2019-0260

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*(2), 255–276. https://doi.org/10.1111/ijmr.12129

Nip, K., Wang, Z., & Wang, Z. (2021). Assortment Optimization under a Single Transition Choice Model. *Production and Operations Management*, *30*(7), 2122–2142. https://doi.org/10.1111/poms.13358

OECD. (2020, October). *E-commerce in the time of COVID-19* (OECD Policy Responses to Coronavirus (COVID-19)). OECD Publishing. Paris. https://doi.org/10.1787/3a2b78e8-en

Phillipson, F., & Van Kempen, E. (2021). Flexibility in Home Delivery by Enabling Time Window Changes. *Proceedings of the 10th International Conference on Operations Research and Enterprise Systems*, 453–458. https://doi.org/10.5220/0010330604530458

Powell, W. B. (2016). Perspectives of approximate dynamic programming. *Annals of Operations Research*, *241*(1-2), 319–356. https://doi.org/10.1007/s10479-012-1077-6

Prokhorchuk, A., Dauwels, J., & Jaillet, P. (2019). Stochastic Dynamic Pricing for Same-Day Delivery Routing. *arXiv preprint*, *arXiv:1912.02946*.

Punakivi, M., & Saranen, J. (2001). Identifying the success factors in e-grocery home delivery. *International Journal of Retail & Distribution Management*, *29*(4), 156–163. https://doi.org/10.1108/09590550110387953

Ramaekers, K., Caris, A., Moons, S., & Van Gils, T. (2018). Using an integrated order picking-vehicle routing problem to study the impact of delivery time windows in e-commerce. *European Transport Research Review*, *10*(2), 56. https://doi.org/10.1186/s12544-018-0333-5

Rodríguez García, M., González Romero, I., Bas, Á. O., & Prado-Prado, J. C. (2022). E-grocery retailing: From value proposition to logistics strategy. *International Journal of Logistics Research and Applications*, *25*(10), 1381–1400. https://doi.org/10.1080/13675567.2021.1900086

Rooderkerk, R., De Leeuw, S., & Hübner, A. (2023). Advancing the marketing-operations interface in omnichannel retail. *Journal of Operations Management*, *69*(2), 188–196. https://doi.org/10.1002/joom.1241

Roth, A. V., & Menor, L. J. (2003). Insights into service operations management: A research agenda. *Production and Operations Management*, *12*(2), 145–164. https://doi.org/10.1111/j.1937-5956.2003.tb00498.x

Schwamberger, J., Fleischmann, M., & Strauss, A. (2023). Feeding the Nation—Dynamic Customer Contacting for E-Fulfillment in Times of Crisis. *Service Science*, *15*(1), 22–40. https://doi.org/10.1287/serv.2022.0304

Siragusa, C., & Tumino, A. (2021). E-grocery: Comparing the environmental impacts of the online and offline purchasing processes. *International Journal of Logistics Research and Applications*, *25*(8), 1164–1190. https://doi.org/10.1080/13675567.2021.1892041

Smilowitz, K. R., & Daganzo, C. F. (2007). Continuum approximation techniques for the design of integrated package distribution systems. *Networks*, *50*(3), 183–196. https://doi.org/10.1002/net.20189

Snoeck, A., Merchán, D., & Winkenbach, M. (2020). Revenue management in last-mile delivery: State-of-the-art and future research directions. *Transportation Research Procedia*, *46*, 109–116. https://doi.org/10.1016/j.trpro.2020.03.170

Soeffker, N., Ulmer, M. W., & Mattfeld, D. C. (2017). On Fairness Aspects of Customer Acceptance Mechanisms in Dynamic Vehicle Routing. In R. O. Lange, N. Kramer, A.-K. Radig, M. Schäfer, & A. Sulzbach (Eds.), *Proceedings of Logistik Management 2017* (pp. 17–24). University of Vienna.

Spliet, R., Dabia, S., & Van Woensel, T. (2018). The Time Window Assignment Vehicle Routing Problem with Time-Dependent Travel Times. *Transportation Science*, *52*(2), 261–276. https://doi.org/10.1287/trsc.2016.0705

Spliet, R., & Desaulniers, G. (2015). The discrete time window assignment vehicle routing problem. *European Journal of Operational Research*, *244*(2), 379–391. https://doi.org/10.1016/j.ejor.2015.01.020

Spliet, R., & Gabor, A. F. (2015). The Time Window Assignment Vehicle Routing Problem. *Transportation Science*, *49*(4), 721–731. https://doi.org/10.1287/trsc.2013.0510

Strauss, A., Gülpınar, N., & Zheng, Y. (2021). Dynamic pricing of flexible time slots for attended home delivery. *European Journal of Operational Research*, *294*(3), 1022–1041. https://doi.org/10.1016/j.ejor.2020.03.007

Strauss, A. K., Klein, R., & Steinhardt, C. (2018). A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research*, *271*(2), 375–387. https://doi.org/10.1016/j.ejor.2018.01.011

Stroh, A. M., Erera, A. L., & Toriello, A. (2022). Tactical Design of Same-Day Delivery Systems. *Management Science*, *68*(5), 3444–3463. https://doi.org/10.1287/mnsc.2021.4041

Talluri, K., & Van Ryzin, G. (2004). Revenue Management Under a General Discrete Choice Model of Consumer Behavior. *Management Science*, *50*(1), 15–33. https://doi.org/10.1287/mnsc.1030.0147

Talluri, K. T., & Van Ryzin, G. (2004). *The theory and practice of revenue management*. Kluwer Academic Publishers.

Truden, C., Maier, K., Jellen, A., & Hungerländer, P. (2022). Computational Approaches for Grocery Home Delivery Services. *Algorithms*, *15*(4), 125. https://doi.org/10.3390/a15040125

Ulmer, M. (2017). Delivery Deadlines in Same-Day Delivery. *Logistics Research*, *10*(3), 1–15. https://doi.org/10.23773/2017_3

Ulmer, M. W. (2019). Anticipation versus reactive reoptimization for dynamic vehicle routing with stochastic requests. *Networks*, *73*(3), 277–291. https://doi.org/10.1002/net.21861

Ulmer, M. W. (2020). Dynamic Pricing and Routing for Same-Day Delivery. *Transportation Science*, *54*(4), 1016–1033. https://doi.org/10.1287/trsc.2019.0958

van Brouwershaven, R. (2020). *Exterior design of a bigger, faster, stronger last-mile delivery vehicle for Picnic Technologies* [Master's thesis, TU Delft].

van der Hagen, L., Agatz, N., Spliet, R., Visser, T. R., & Kok, L. (2022). Machine Learning–Based Feasibility Checks for Dynamic Time Slot Management. *Transportation Science*, *58*(1), 1–278. https://doi.org/10.1287/trsc.2022.1183

Vinsensius, A., Wang, Y., Chew, E. P., & Lee, L. H. (2020). Dynamic Incentive Mechanism for Delivery Slot Management in E-Commerce Attended Home Delivery. *Transportation Science*, *54*(3), 567–587. https://doi.org/10.1287/trsc.2019.0953

Visser, T. R., Agatz, N., & Spliet, R. (2024). Managing Concurrent Interactions in Online Time Slot Booking Systems for Attended Home Delivery. *Transportation Science*, *58*(5), 1056–1075. https://doi.org/10.1287/trsc.2022.0445

Visser, T. R., & Savelsbergh, M. W. P. (2019). Strategic Time Slot Management: A Priori Routing for Online Grocery Retailing. *Working Paper*, *EI2019-04*.

Wagner, L., Pinto, C., & Amorim, P. (2021). On the Value of Subscription Models for Online Grocery Retail. *European Journal of Operational Research*, *294*(3), 874–894. https://doi.org/10.1016/j.ejor.2020.05.011

Wang, R. (2012). Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters*, *40*(6), 492–497. https://doi.org/10.1016/j.orl.2012.08.003

Wang, Y., & Shen, Z.-J. M. (2021). Constrained Assortment Optimization Problem under the Multilevel Nested Logit Model. *Production and Operations Management*, *30*(10), 3467–3480. https://doi.org/10.1111/poms.13443

Waßmuth, K., Agatz, N., & Fleischmann, M. (2025). Evaluating Time Window Assortments for Grocery Home Delivery. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.5641010

Waßmuth, K., Köhler, C., Agatz, N., & Fleischmann, M. (2023). Demand management for attended home delivery – A literature review. *European Journal of Operational Research*, *311*(3), 801–815. https://doi.org/10.1016/j.ejor.2023.01.056

Wei, K., Vaze, V., & Jacquillat, A. (2020). Airline Timetable Development and Fleet Assignment Incorporating Passenger Choice. *Transportation Science*, *54*(1), 139–163. https://doi.org/10.1287/trsc.2019.0924

Wilson-Jeanselme, M., & Reynolds, J. (2006). Understanding shoppers' expectations of online grocery retailing. *International Journal of Retail & Distribution Management*, *34*(7), 529–540. https://doi.org/10.1108/09590550610673608

Yang, X., & Strauss, A. K. (2017). An approximate dynamic programming approach to attended home delivery management. *European Journal of Operational Research*, *263*(3), 935–945. https://doi.org/10.1016/j.ejor.2017.06.034

Yang, X., Strauss, A. K., Currie, C. S. M., & Eglese, R. (2016). Choice-Based Demand Management and Vehicle Routing in E-Fulfillment. *Transportation Science*, *50*(2), 473–488. https://doi.org/10.1287/trsc.2014.0549

Yildiz, B., & Savelsbergh, M. (2020). Pricing for delivery time flexibility. *Transportation Research Part B: Methodological*, *133*, 230–256. https://doi.org/10.1016/j.trb.2020.01.004

Zhang, J., Deng, L., Liu, H., & Cheng, T. (2022). Which strategy is better for managing multi-product demand uncertainty: Inventory substitution or probabilistic selling? *European Journal of Operational Research*, *302*(1), 79–95. https://doi.org/10.1016/j.ejor.2021.11.055

Short CV

Katrin Waßmuth

Professional Experience

Since 2025	Research Assistant Chair of Data Analytics & Statistics Bundeswehr University Munich, Germany
2020 – 2025	Research Assistant Chair of Supply Chain Management University of Mannheim, Germany

Education

2020 – 2025	Doctoral Studies in Business Administration (Dr. rer. pol.) University of Mannheim, Germany
2023	Visiting Researcher Erasmus University Rotterdam, Netherlands
2018 – 2020	Mannheim Master in Management (M.Sc.) University of Mannheim, Germany
2016	Exchange Semester University of Windsor, Canada
2014 – 2018	Mathematics in Business and Economics (B.Sc.) University of Mannheim, Germany