Demand Management for E-Groceries

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

submitted by

Jonas Schwamberger
Mannheim

Dean: *Joachim Lutz*
Referent: *Prof. Dr. Moritz Fleischmann*
Correferent: *Prof. Dr. Arne Strauss*

Day of oral examination: 05.05.2023

Acknowledgements

This dissertation would not have been possible without the numerous support I received over the past years. At this point, I would like to express my deepest gratitude to all people who accompanied me along this journey.

First and foremost, I would like to thank my doctoral advisor Moritz Fleischmann who facilitated these last years and the numerous wonderful memories. I am very grateful to him for putting his trust in me and always being available for discussing my research. It was his valuable feedback that made this successful dissertation possible. Likewise, I am very grateful to co-supervisor Arne Strauss who supported me in all projects and always provided me with his helpful comments.

My current and former colleagues at the Area Operations Management at the University of Mannheim also deserve special gratitude. In particular, I would like to thank my fellow doctoral students Luca Biscaro, Hilal Müge Cakan, Christoph Löffel, Sebastian Müller, Hossein Nikpayam, Maryam Nouri Roozbahani, Pardis Sahraei, Kilian Seifried, Sebastian Szilvas, Katrin Waßmuth, and our secretaries Helke Naujok, Julia Peemöller, and Ruth Pfitzmann. It was you who have made the past years memorable.

Finally, I am incredibly grateful to my friends and family. I would especially like to highlight the support I received from Geneviève Schmidthermes, Marcus Kauke, and Jennifer Pignol. They have always believed in me, supported me, and encouraged me to pursue my doctorate. Without all their unconditional help and support, this dissertation would not have been possible!

Summary

The rapidly emerging trend of purchasing products online has not spared groceries. Since groceries are fresh and partly refrigerated goods, delivery can only be successfully completed when customers are at home to receive their deliveries. This complicates the fulfillment of customer orders as e-grocers and customers have to agree on mutual time windows for delivery. This dissertation is concerned with demand management for e-grocers. Chapter 2 of this dissertation addresses an e-grocer's time slot offering problem from a customer choice perspective. To this end, the time slot offering problem is modeled as an online assortment optimization problem using a consider-then-choose customer choice model. An estimation procedure for the consider-then-choose customer choice model based on historical transaction data is proposed which incorporates a predefined cluster structure. We show that the estimation procedure can identify the underlying customer choice behavior and that the underlying cluster structure can be efficiently exploited in the online time slot offering problem. Chapter 3 focuses on an acute crisis situation of an e-grocer in which demand outstrips delivery capacity by a wide margin, such as during the COVID-19 pandemic. In this chapter, we examine how the demand management lever of proactive customer contacting can be utilized to manage the customer arrival process. We propose a contacting scheme for the fulfillment channels Click&Collect and Attended Home Delivery and show that proactive customer contacting can be used to prioritize vulnerable customers at the expense of a relatively small decline in overall revenue. Chapter 4 explores the lever of proactive customer contacting in post-crisis situations to balance demand disparity over several days. Here, proactive customer contacting is used as a contingency tool when it is anticipated that not enough customers will emerge via the classical booking process. Compared to Chapter 3, the customer arrival process not only consists of contacted customers but also comprises non-contacted customers. This leads to a cannibalization effect that must be managed. We demonstrate the merit of proactive customer contacting and compare different contacting strategies in various settings. In particular, we reveal what drives the success of proactive customer contacting.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AHD	Attended Home Delivery
C&C	Click&Collect
CNL	Cross-Nested Logit
DP	Dynamic Programming
EDV	Elderly, Disabled & Vulnerable
FCFS	First Come First Served
GAM	Generalized Attraction Model
MAD	Mean Absolute Deviation
MMNL	Mixed-MNL
MNL	Multinomial Logit
NL	Nested Logit
RCS	Random Consideration Set
RMSE	Root Mean Squared Error
STSP	Selective Traveling Salesman Problem
TOP	Team Orienteering Problem
TOPTW	Team Orienteering Problem with Time Windows

Chapter 1.

Introduction

October 29^{th}, 1969, is a very important but yet relatively unknown date in recent history. On this day, the first message was sent over a network now known as the Internet (Kleinrock 2010). Since then, the development of this technology has fundamentally changed the daily lives of many people, as the Internet has had a lasting impact on virtually every sphere of life. With the advent of e-commerce, for instance, shopping behavior has experienced a radical transformation, with many products being purchased online (Statista 2022). This development has not spared groceries (see, e.g., REWE or Sainsbury's). Compared to other product classes, ordering groceries online involves several additional challenges. Since most products are fresh and some groceries need to be refrigerated, one issue is that customers need to be at home to receive the groceries when their orders are delivered. These attended home deliveries complicate the fulfillment process as e-grocers and customers must mutually agree on time windows to successfully complete deliveries. To succeed in the online market, e-grocers must carefully govern their operations and fulfillment processes. One tool for guiding these processes is customer demand management which is the central topic of this dissertation.

Chapter 2 of this dissertation investigates the time slot offering problem, i.e., the question of which time slots to offer to arriving customers. This problem is of tremendous importance because e-grocers operate in a particularly challenging market situation with low profit margins, and the selected time windows have an immense impact on the efficacy of the offered service. So far, the time slot offering problem has mostly been studied with a focus on the routing aspect rather than the customer choice component (Wassmuth et al. 2023). In this dissertation, we dedicate Chapter 2 to examine this problem from a customer choice perspective. We model the time slot offering problem as an online assortment optimization problem using a consider-then-choose customer choice model. We propose a method how to estimate the consider-then-choose choice model from historical transaction data that incorporates a predefined cluster structure. This cluster structure is then exploited in the online time slot offering problem. In a numerical study, we evaluate the estimation procedure and the online time slot offering problem. In particular, we show that the proposed estimation method of the consider-then-choose choice model works and is able to identify the underlying customer choice behavior for many time slots over multiple days. For the time slot offering problem, we demonstrate that the chosen assortment optimization approach can efficiently exploit the underlying cluster structure and leads to fast online computation times. This is particularly important in the considered online context.

The increasing trend of purchasing products online was further amplified by the COVID-19 crisis (Mintel 2021, McKinsey 2020). This rapid increase in demand exceeded the existing capacities, leading to e-grocers no longer being able to serve all customer requests and booking systems being unable to cope with the demand onslaught (CNA 2020, Telegraph 2020, Vox 2020). In this situation, some e-grocers began to proactively reach out to customers with the aim of serving as many high-priority customers as possible. In Chapter 3, we focus on such a crisis situation and present approaches how to coordinate proactive customer contacting for both fulfillment channels Click&Collect (C&C) and Attended Home Delivery (AHD). In a realistic numerical study based on the region of London, UK, we evaluate proactive customer contacting compared to the traditional reactive order intake and derive several pioneering managerial insights. We observe that proactive customer contacting is a valuable demand management tool to prioritize vulnerable customer groups and show that especially high customer response rates, as prevalent in crisis situations, benefit the contacting policy. Moreover, we demonstrate that the benefit of adding an additional contacting possibility decreases quickly and that the trade-off between prioritizing vulnerable customer groups and revenue is relatively mild.

Based on the insights gained in crisis situations, naturally the question arises of how proactive customer contacting can be leveraged in post-crisis times. Chapter 4 addresses this issue and explores whether proactive customer contacting can be used as a contingency tool to balance the demand disparity over multiple days. In this setting, proactive customer contacting is activated if the e-grocer anticipates that not enough customer inquiries will emerge via the normal booking process. Here, proactive customer contacting serves as a demand management lever to shift future demand to an earlier point in time. Compared to Chapter 3, the arrival process does not only consist of contacted customers but now also comprises contacted and non-contacted customers. This leads to a cannibalization effect that must be managed. In a numerical study, we compare different contacting strategies and explore the value of proactive customer contacting in post-crisis times. In particular, we observe that it is not always optimal to contact the marginal best customers. Instead, the optimal contacting decision depends on the interrelationship of customers available for contacting. In addition, we demonstrate that the benefit of contacting depends to a great extent on the overall system utilization, the utilization of the preliminary route, and the response probability of the contacted customers.

Chapter 2.

Tractable Time Slot Assortment Optimization in Attended Home Delivery under Consider-Then-Choose Customer Choice[1]

with Moritz Fleischmann and Arne Strauss

[1]The research presented in this chapter is based on a paper entitled "Tractable Time Slot Assortment Optimization in Attended Home Delivery under Consider-Then-Choose Customer Choice", coauthered with Moritz Fleischmann and Arne Strauss (Schwamberger et al. 2023).

Abstract

In attended home delivery, the delivery time slot offering problem has a significant impact on the efficiency of the retailer providing this service. Since the decision on which time slots to offer is typically made online during the customer booking process, this problem requires a real-time solution. However, solving this problem is complex since most companies offer multiple time slots over multiple days and must account for customer preferences. To adequately reflect customer choice, an appropriate customer choice model must be adopted. We employ the consider-then-choose choice model, which in the empirical literature has been shown to reflect the general underlying customer choice behavior well. We address the time slot offering problem from a customer choice perspective. In particular, we propose a time slot assortment optimization model that exploits a customer cluster structure underlying the consider-then-choose choice model to solve the time slot offering problem in real-time. In addition, we propose a method for estimating the consider-then-choose choice model with such a cluster structure from historical transaction data. We evaluate this estimation method and the proposed time slot offering model in a numerical study and demonstrate that the estimation procedure can extract the underlying choice behavior and that the time slot offering problem can be solved in real-time for realistic problem sizes of an e-grocer.

2.1. Introduction

In recent years, e-grocers have been experiencing strong growth in demand. In particular, the COVID-19 pandemic has contributed to the growing popularity of e-grocers' attended home delivery service (Schwamberger et al. 2022b). In this service, customers usually specify their delivery address and shopping basket before selecting one of the offered time slots in which they prefer to obtain their delivery. E-grocers operate in a particularly challenging market situation. Due to fierce competition, profit margins are low, requiring efficient fulfillment of customer orders. Since customers' time slot decisions significantly affect the efficiency of the delivery tours and thus the efficiency of the offered service, they have a major impact on the profitability of online grocery retailers.

The time slot offering problem, i.e., deciding which time slots to offer to an arriving customer, consists of two core elements: the customer choice and the routing component. The literature has mainly focused on the routing component of the problem, whereas the customer choice element has received less attention and is typically represented using simple parametric choice models (Ehmke and Campbell 2014, Lang et al. 2021a, Wassmuth et al. 2023). However, adequately mapping customer behavior over multiple days is nearly impossible with these simple approaches and their underlying strong assumptions. The time slot offering problem is further complicated by the fact that it is usually solved in real-time since the profitability of offering a given time slot to an arriving customer depends on already accepted customers and customer-specific information (such as location). Hence, the problem must be solved quickly because offered time slots must be displayed to customers immediately. Especially given many time windows over multiple days, solving this problem in real-time becomes a significant challenge.

In this paper, we focus on the time slot offering problem from a *customer choice* perspective and contribute to the literature by addressing two interrelated problems. First, we address the problem of determining the optimal time slot assortment, given the customer preferences for the individual time slots. For this purpose, we interpret the time slot offering problem as an online assortment optimization problem and propose a tractable optimization formulation to solve this problem for a specific non-parametric choice model. Second, we investigate how to estimate the customer preferences for the different time slots from historical transaction data.

To adequately reflect customer preferences and capture the resulting behavior, we make use of the consider-then-choose customer choice model. Empirical literature from the field of marketing and psychology (Hauser 2014, Liu and Arora 2011, Aouad et al. 2020) yields strong indications that the customer's actual decision-making process comprises these two steps. In the first step, all available items are screened, and those that are considered for purchase are identified. In the second step, the considered items are ranked so that a purchase decision can be made. In our context of selecting delivery time slots, it seems reasonable that customers choose in a similar manner. First, they only consider those time slots for which they are at home such that a delivery can successfully be completed (i.e., customers form their consideration set). Subsequently, customers rank the time slots and choose the slot that seems most convenient to them. In our analysis, we focus on this empirically established non-parametric customer choice model.

Regarding the individual customers and their time window choice process, it appears natural that different customers consider identical time windows in the first step. For example, most customers will only consider time slots in the evening since that is the only time they will be home to receive their deliveries. We exploit this underlying market structure by clustering customers based on their time slot consideration set. Even if customers consider identical time slots, they may still have different preferences regarding the considered time slots.

Our contribution to the literature is a tractable model formulation for an e-grocer's online time slot offering problem over multiple days using a consider-then-choose customer choice model with an underlying cluster structure. In addition, we propose a procedure for estimating the consideration sets and preference lists of the consider-then-choose model based on historical transaction data, while respecting the assumed cluster structure.

This paper is organized as follows: In Section 2.2, we review and discuss the related literature. In Section 2.3, we define the general problem setting. In Section 2.4 and Section 2.5, we discuss the assortment optimization and customer choice estimation process, respectively. Section 2.6 describes the results of our numerical study, and Section 2.7 presents our conclusions.

2.2. Related Literature

The work presented in this paper lies at the intersection of two literature streams: (i) demand management in the context of e-groceries and (ii) customer choice behavior and assortment optimization.

2.2.1. Demand Management in the Context of E-Groceries

In recent years, demand management has gained increasing importance in logistics environments (Agatz et al. 2011, Yang et al. 2016, Yildiz and Savelsbergh 2020, Paul et al. 2019), especially in the field of e-groceries. Agatz et al. (2013) identified time slot allocation and time slot pricing as the two main levers of demand management for AHD. These decisions are made on a strategic, tactical or operational level. A detailed overview of recent work in the field of demand management for e-grocers can be found in Wassmuth et al. (2023). Studies on operational time slot offering (Campbell and Savelsbergh 2005, Ehmke and Campbell 2014, Mackert 2019) are closest to our work. These papers mainly consider the problem from a routing perspective rather than from a customer choice perspective. In most of the works, customer choice is modeled by an independent demand model (Campbell and Savelsbergh 2005, Ehmke and Campbell 2014) or by a simple parametric customer choice model, such as the multinomial logit (MNL) model (Lang et al. 2021a,b) or the generalized attraction model (GAM) (Mackert 2019). None of these studies, however, include a non-parametric customer choice model that does not require stringent assumptions, is capable of capturing customer choice behavior over multiple days, and can be solved in real-time. We contribute an estimation and optimization approach for a consider-then-choose customer choice model for the time slot offering problem to this stream of literature. To the best of our knowledge, we are the first to study the time slot offering problem from a customer choice perspective by incorporating the consider-then-choose choice model in this setting. For a further overview of managerial planning tasks, we recommend Agatz et al. (2008a). Recent advantages of integrating demand management and vehicle routing can be found in the review of Fleckenstein et al. (2023).

2.2.2. Customer Choice Modeling and Assortment Optimization

The focus of modeling customer choice behavior in assortment optimization has evolved from independent demand models to choice-based models that capture the effect of substitutions among different products. The literature on discrete customer choice models can be classified into parametric and non-parametric models (for a more detailed classification, see Strauss et al. (2018)). Because our approach is based on a non-parametric customer choice model, we only provide a brief overview of parametric models and then describe non-parametric models in greater detail.

Parametric customer choice models are rooted in random utility theory and assume that products are associated with certain utilities. The underlying concept is that the customer choice behavior is a maximization mechanism (Strauss et al. 2018). Some famous examples of parametric customer choice models are logit models, for instance the MNL model, the nested logit (NL) model and the mixed-MNL (MMNL) model. Although parametric customer choice models provide an initial attempt to represent customer preferences, these models are based on the strong assumption that the customer choice behavior can be captured in a certain functional form using the utilities associated with the products. While capturing customer choice in a functional form facilitates the application of the model, it limits the flexibility of choice modeling.

Most recently, non-parametric customer choice models have received more attention. These models allow a more flexible representation of choice behavior. One example of a non-parametric customer choice model is the rank-based choice model. This model assumes that a consumer ranks all products and the no-purchase option in a certain order and chooses the highest-ranked available option (Van Ryzin and Vulcano 2015, Farias et al. 2013).

A more sophisticated non-parametric ranked-based customer choice model is the consider-then-choose model, which is considered in this paper. According to Hauser (2014) and Liu and Arora (2011), there are strong indications that this model provides a good representation of a customer's decision-making process. The general consider-then-choose model for assortment planning has been investigated in several papers (e.g., Aouad et al. (2020), Bertsimas and Mišić (2019)). As in our study, some works have focused on special cases of the consider-then-choose model. For instance, Paul et al. (2016) investigate the consider-then-choose model with "choosy" customers, i.e., a customer considers purchasing one of two substitutable products. Feldman et al. (2019)

similarly investigate the consider-then-choose choice model with a limited number of products in the consideration set. Jagabathula and Rusmevichientong (2017) focus on another special case of the consider-then-choose model with a particular policy for forming the consideration sets. They assume that customers follow a two-stage choice process in which they first consider all products with a price less than a certain threshold and then choose the most preferred option. Another special case of the consider-then-choose model is the random consideration set (RCS) model considered by Manzini and Mariotti (2014). This model assumes an underlying unique preference order of the products and presumes that the products have certain attention probabilities that are used to determine the consideration sets. Gallego and Li (2017) also investigate the RCS model. As in the aforementioned papers, we focus on a special case of the consider-then-choose customer choice model by assuming an underlying customer cluster structure.

As generally no structural properties are assumed in non-parametric customer choice models, the estimation of these models becomes more challenging than parametric models. However, when assuming a special structure of the choice model, the estimation procedure can be tailored to this. For example, Jagabathula et al. (2019) propose a framework to effectively estimate the model with a unique preference order over products and infer consideration sets based on machine learning techniques. We do not impose a single preference order over all products but include a cluster structure in our estimation procedure. Our estimation process uses ideas from Van Ryzin and Vulcano (2015) and Van Ryzin and Vulcano (2017). In particular, Van Ryzin and Vulcano (2015) propose a general approach for estimating customer preferences for suitable products from historical transaction data and assume an underlying rank-based discrete choice model of preferences with a Bernoulli customer arrival process. In our work, we apply these concepts to the consider-then-choose setting and incorporate an underlying customer cluster structure.

In summary, the consider-then-choose customer choice estimation and the time slot optimization approach presented in this paper draw inspiration from the different studies outlined above. However, there is no previous research that has addressed the time slot offering problem of an e-grocer under consider-then-choose customer choice behavior. Our paper contributes to the literature by proposing a tractable online time slot offering model and by introducing an estimation approach that is compatible with an underlying cluster structure.

2.3. Problem Setting

We consider an e-grocer's online time slot offering problem. More precisely, the e-grocer needs to decide which time slots to offer to an arriving customer with known location and shopping basket. We consider this problem as a dynamic assortment optimization problem. In our setting, the time slots represent the "products" that can potentially be offered. We assume that there are $N \in \mathbb{N}$ different products and the retailer must decide which of these products to include in the assortment $A \subset \{1, \ldots, N\}$. Each time slot $i \in \mathcal{N} = \{1, \ldots, N\}$ is associated with a profit margin r_i. To be consistent with the assortment optimization literature, we denote these margins as "revenues". We assume that these profit margins are known for an arriving customer. For example, these margins can be estimated using the obtained revenue (including delivery fee), the cost of fulfilling the customer's request, e.g., estimated routing costs, and the potential opportunity cost. Here, the e-fulfillment literature focusing on the routing component of the problem can be used. For instance, the procedures proposed in Yang et al. (2016) or Klein et al. (2018) can be used to approximate these costs and thus the profit margins r_i. Calculating the profit margins is not the subject of this study. We rather focus on determining which time slots to offer, given the different margins and the estimated customer choice behavior.

As outlined above, we model the customer choice behavior in this assortment problem by a consider-then-choose choice model. This model assumes $L \in \mathbb{N}$ customer segments. Each customer segment $l \in \{1, \ldots, L\}$ is associated with a weight $\lambda_l \in [0, 1]$. These weights can be interpreted as different arrival probabilities or different segment sizes. A customer segment is characterized by a consideration set, i.e., the set of products that is considered for purchase, and by a preference list, i.e., the sequence in which the products are ranked. The consideration set of customer segment l is denoted by $\emptyset \neq C_l \subset \{1, \ldots, N\}$. The preferences of the individual customer segments for products are recorded in preference lists of the considered products. The ranking of products for customer segment l is denoted by a permutation σ_l that describes the unique preference order of the products in the consideration set. If there are several products $a \in A \cap C_l$ under consideration contained in the offered assortment, customer segment l purchases the product for which $\sigma_l(a)$ is lowest, i.e., most preferred. The revenue $R_l(A)$ of an offered assortment A obtained from customer segment l can then be calculated as

$$R_l(A) = \begin{cases} r_{a^*} & a^* = \arg\min_{a \in A \cap C_l} \sigma_l(a) \\ 0 & A = \emptyset \text{ or } A \cap C_l = \emptyset \end{cases}.$$

In our application context, it occurs naturally that different customer segments consider similar time slots in the first step. For instance, most employees are home only in the evening and thus consider evening time slots only. To model this, we assume that the customer segments can be divided into clusters, based on their consideration sets. We assume that we can assign a customer segment to a given cluster $c \in \mathcal{C}$ that is represented by the set of products considered in the cluster, i.e., $c \subset \mathcal{N}$. Hence, the consideration set of segment l belonging to a customer cluster c is a subset of the time slots considered by the cluster, i.e., $C_l \subset c$. We use this notation to formulate the online time slot assortment problem in Section 2.4, describe the corresponding estimation process in Section 2.5, and evaluate the estimation and optimization approach for different underlying cluster structures in Section 2.6.

2.4. Assortment Optimization

The goal of the general assortment optimization problem for a given set of products $\mathcal{N}$ is to determine the subset of products $A \subset \mathcal{N}$ that maximizes the obtained revenues: $\max_{A \subset \mathcal{N}} R(A)$. There are several ways to model this optimization problem. Given formulations differ in particular in their decision variables. For instance, some model formulations decide on product-level which products to offer (Aouad et al. 2020), while others decide on subset-level of the entire assortment (Meissner et al. 2013). Although these model formulations are very intuitive and accessible, they do not exploit any properties of the underlying customer choice model and become very large (in terms of the number of decision variables and number of constraints) and complex to solve. In our real-time application of offering time slots over multiple days, these model formulations do not work due to relatively long computation times (i.e., more than 1 second). In contrast to these general model formulations, we seek to leverage the underlying structure of the consider-then-choose customer choice model. More precisely, our approach to model the time slot offering problem is based on the idea of determining which subsets of the consideration sets to offer to the various customer segments (Talluri 2010, Meissner et al. 2013). This is captured by decision variables $x^l_{S_l} \in \{0, 1\}$ that indicate whether subset S_l of consideration set C_l is offered to customer segment l. When using these decision variables, it becomes necessary to ensure consistency of offered time slots across customer segments. This can be achieved by appropriate constraints (see below). Furthermore, note that the total offered assortment is only implicitly included in this model formulation. It can be obtained by combining the activated decision variables of the individual customer segments:

$$A = \bigcup_{l \in \{1,\dots,L\},\, S_l \subset C_l \,:\, x^l_{S_l}=1} S_l.$$

The assortment optimization in terms of these decision variables can be formulated as follows:

$$\max \sum_{l=1}^{L} \lambda_l \sum_{S_l \subset C_l} R_l(S^l) x^l_{S_l} \tag{2.1}$$

$$\sum_{S_l \subset C_l} x^l_{S_l} = 1 \qquad \forall\, l \in \{1, \ldots, L\} \tag{2.2}$$

$$\sum_{S_l \supset S_{lm}} x^l_{S_l} = \sum_{S_m \supset S_{lm}} x^m_{S_m} \quad \forall\, \emptyset \neq S_{lm} \subset C_l \cap C_m,\, l, m \in \{1, \ldots, L\} : C_l \cap C_m \neq \emptyset \tag{2.3}$$

$$x^l_{S_l} \in \{0, 1\} \qquad \forall\, l \in \{1, \ldots, L\},\; S_l \subset C_l. \tag{2.4}$$

The objective function (2.1) maximizes the weighted total sum of revenues received from the various customer segments. The first set of constraints (2.2) ensures that every customer segment l obtains exactly one subset of potential products as an individual assortment. The next set of constraints (2.3), the so-called product cut constraints, were first introduced and investigated in Meissner et al. (2013) and Strauss and Talluri (2017). These constraints ensure consistency of the product offering across customer segments. Specifically, if a set of products $\emptyset \neq S \in C_l \cap C_m$ is offered to customer segment C_l, then it must also be offered to customer segment C_m since segment memberships are unobservable. Constraints (2.4) define the domains of the decision variables.

Thus far, we have not imposed any structural requirements on the customer segments. In our application context, we assume that some customers have similar consideration sets and can be divided into clusters, based on their considered time slots (Section 2.3). Exploiting such a cluster structure $\mathcal{C}$ is the central idea of our approach. We introduce decision variables $x^c_{S_c} \in \{0, 1\}$ that determine which subset of products $S_c \subset c$ to offer to customer cluster $c \in \mathcal{C}$. The revenues obtained from cluster c when offering product subset A can be calculated *before* starting the online optimization process as the probability-weighted sum of all individual revenues from all contained customer segments l, i.e.,

$$R_c(A) = \sum_{l \in c} \lambda_l R_l(A).$$

Depending on the cluster structure, this aggregation of customer segments can improve the efficiency of the real-time optimization problem since this data can be preprocessed. For a given set of clusters, the model formulation (2.1)-(2.4) can thus be equivalently formulated as

$$\max \sum_{c\in\mathcal{C}} R_c(S^c)x^c_{S_c} \tag{2.5}$$

$$\sum_{S_c\subset c} x^c_{S_c} = 1 \qquad \forall\, c \in \mathcal{C} \tag{2.6}$$

$$\sum_{S_c\supset S_{cd}} x^c_{S_c} = \sum_{S_d\supset S_{cd}} x^d_{S_d} \qquad \forall\, \emptyset \neq S_{cd} \subset c\cap d,\ c,d\in\mathcal{C} : c\cap d \neq \emptyset \tag{2.7}$$

$$x^c_{S_c} \in \{0,1\} \qquad \forall\, c\in\mathcal{C},\ S_c \subset c. \tag{2.8}$$

This formulation is similar to that in (2.1)-(2.4) but considers decision variables for customer clusters rather than for segments. The objective (2.5) maximizes the total weights received by all clusters, constraints (2.6) ensure that each cluster obtains one subset of potential products as an individual assortment, and constraints (2.7) guarantee that products offered to one cluster have also to be offered to all other clusters if these products are considered. The problem formulation based on customer clusters captures only indirectly which customer segment chooses which product, namely in the achieved revenues $R_c(S^c)$ of the different subsets. Depending on the cluster structure, this simplifies and accelerates the online optimization procedure because the revenues $R_c(S^c)$ can be precomputed. In particular, the model formulation (2.5) - (2.8) benefits from a low number of clusters and from a small cluster size because fewer decision variables need to be introduced and fewer coupling constraints need to be considered. We evaluate the impact of the underlying cluster structure in a numerical study in Section 2.6.2.1.

2.5. Estimation of the Customer Choice Model

To determine the online assortment, the consider-then-choose customer choice model consisting of consideration sets, preference lists and the respective weights of the individual customer segments must be given. In this section, we propose an estimation procedure to obtain the customer choice model from historical transaction data while fitting it into a given customer cluster structure $\mathcal{C}$. How to design an appropriate cluster structure $\mathcal{C}$ is discussed in Section 2.6.2.2. The historical selling horizon consists of T periods reflected in the sequence of tuples $((j_1, S_1), \ldots, (j_t, S_t), \ldots, (j_T, S_T))$ in which S_t describes the offered assortment and j_t the chosen product in period $t \in \{1, \ldots, T\}$. For our context of offering time slots, we assume that we can observe periods with arriving customers that do not select any of the time slots. To reflect this, we include the no-purchase option as an artificial product "0" in all offered assortments. To obtain a customer choice model retaining the desired cluster structure that can be leveraged in online assortment optimization, we embed the estimation into a given customer cluster structure $\mathcal{C}$. Based on this cluster structure, we generate potential customer segments in a first step. To this end, we use all permutations $\sigma_i \in \mathcal{P}(c)$ of the products considered in a cluster $c \in \mathcal{C}$ as individual preference lists for the customer segments. This ensures that each customer segment belongs to a customer cluster and only considers a subset of products that is considered by the respective cluster. These are the customer segments available for the following estimation procedure. We use the maximum likelihood approach to determine the probability distribution λ over these customer segments. For this purpose, we introduce a cluster-specific compatible set definition, which is based on the idea introduced by Van Ryzin and Vulcano (2015) but is tailored to the specific customer cluster structure. We define the *compatible set of customer segments from cluster $c \in \mathcal{C}$ for period t* by

$$\mathcal{M}(t, c) = \{\sigma_i \in \mathcal{P}(c) \,|\, \sigma_i(j_t) < \sigma_i(k), \forall\, k \in S_t \cap c, k \neq j_t\},$$

where $\mathcal{P}(c)$ denotes the set of permutations of the products and thus all customer segments in cluster c. The set $\mathcal{M}(t, c)$ contains all customer segments within customer cluster c that are compatible with the decision in period t, i.e., for which the selected product j_t in period t is the most preferred of the available products S_t. The goal is to estimate a probability distribution λ composed of all likelihoods λ_i^c of all customer segments $i \in \mathcal{P}(c)$ for all customer clusters $c \in \mathcal{C}$. To formulate the estimation model, we must determine the probability that a random arriving customer chooses product j_t given offer set S_t. These probabilities can be calculated as

$$\mathbb{P}(j_t \mid S_t) = \sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{M}(t,c)} \lambda_i^c.$$

Using these probabilities, the likelihood function $L(\lambda)$ for a distribution λ of the past periods is given by

$$L(\lambda) = \prod_{t=1}^{T} \mathbb{P}(j_t \mid S_t) = \prod_{t=1}^{T} \left(\sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{M}(t,c)} \lambda_i^c \right).$$

This results in the log-likelihood function $\mathcal{L}$

$$\mathcal{L}(\lambda) = \sum_{t=1}^{T} \log \left(\sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{M}(t,c)} \lambda_i^c \right).$$

Using this notation, we formulate the estimation model for the probability distribution λ as a concave nonlinear maximization problem:

$$\max \sum_{t=1}^{T} \log \left(\sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{M}(t,c)} \lambda_i^c \right) \tag{2.9}$$

$$\sum_{c \in \mathcal{C}} \sum_{i \in \mathcal{P}(c)} \lambda_i^c = 1 \tag{2.10}$$

$$0 \leq \lambda_i^c \leq 1 \qquad \forall\, i \in \mathcal{P}(c),\, c \in \mathcal{C}. \tag{2.11}$$

The objective function (2.9) maximizes the log-likelihood function. The constraint (2.10) ensures that the probability distribution over all customer segments and all customer clusters sums to 1, whereas constraints (2.11) guarantee the non-negativity of the distribution.

2.6. Numerical Study

In this section, we present the numerical results of both our proposed estimation procedure for the consider-then-choose customer choice model with a given cluster structure and the performance of our optimization approach. The aim of this section is to show that (i) our estimation approach can identify customer choice behavior for many time slots over multiple days and (ii) our optimization procedure can solve the time slot offering problem of an e-grocer in real-time. All computations were conducted using an HP 440 G5 with a 2.5 GHz Intel Core i5 processor (two cores) using either GUROBI (for linear problems, e.g., formulation (2.5)-(2.8)) or KNITRO (for nonlinear problems, e.g., (2.9)-(2.11)).

2.6.1. Data and Parameters

Before presenting the results of the estimation and optimization procedure, we explain the relevant parameters that we used in this numerical study if not indicated otherwise. We focus on an e-grocer that operates its service from Monday through Saturday and offers one-hour time slots between 6:00 and 20:00, resulting in 14 time slots per day and a total of 84 time slots per week. We assume that the profit margins of the time slots are given and can be calculated using approximation procedures (Yang et al. 2016, Klein et al. 2018). In our experiments, we follow the common approach in the assortment optimization literature and sample these values independently from a log-normal distribution with mean 2 and standard deviation 1 (Aouad et al. 2019).

2.6.2. Estimation

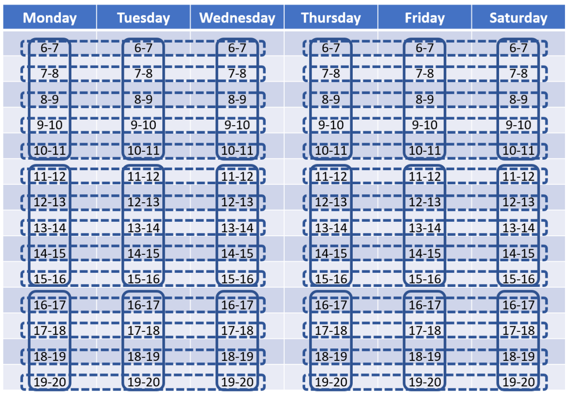
In this subsection, we investigate how well our estimation procedure is able to identify the underlying customer choice behavior over multiple days from historical transaction data. For this purpose, we first evaluate the influence of the predefined cluster structure used in the estimation process (Section 2.6.2.1) and discuss how to choose the predefined cluster structure (Section 2.6.2.2). Subsequently, we modify the underlying customer choice model and explore whether our non-parametric customer choice model can identify accurate purchase probabilities when actual customer behavior is based on a different choice model (Section 2.6.2.3).

To estimate the consider-then-choose customer choice model from historical transaction data and to evaluate how well it reflects the given customer behavior, we use the following two-step process. First, we simulate historical transaction data using a "ground truth" customer choice model. This data serves as training data for the estimation of our consider-then-choose customer choice model. Second, we generate random assortments and compare the predicted purchase probabilities based on our estimated consider-then-choose customer choice model to the underlying true purchase probabilities using the ground truth customer choice model.

In the first step, we simulate historical transaction data using a consider-then-choose ground truth customer choice model, if not indicated otherwise. In this model, we assume that customers use simple screening rules (Gilbride and Allenby 2004, Jagabathula and Rusmevichientong 2017) to obtain their consideration sets. In particular, we assume that an arriving customer screens either a specific day's time slots by time of day or the upcoming delivery days by time slots. Considering one day, customers consider either a time slot in the morning, i.e., between 6:00 and 11:00, a time slot around lunch time, i.e., between 11:00 and 16:00, or a time slot in the late afternoon, i.e., between 16:00 and 20:00. For the selection rule screening the upcoming days by time slots, we assume that customers consider a specific time slot either in the first half of the week, i.e., from Monday-Wednesday, or in the second half of the week, i.e., from Thursday-Saturday. Based on these screening rules, which are visually summarized in Figure 2.1, the cluster structure of the ground truth model is constructed. To generate preference lists based on the consideration sets obtained from these screening rules, we use all possible permutations. As in Aouad et al. (2020), we assume equal weights for all customer segments.

To train our consider-then-choose customer choice model, we use 2,000 historical transaction periods. To determine how well our consider-then-choose model reflects the real underlying customer behavior, we simulate 500 test assortments and compare our results to the true purchase probabilities from the originally chosen choice model. To obtain statistically stable results from our simulations, we indicate the two-sided 95% confidence interval for all results.

Figure 2.1.: Visualization: Ground Truth Cluster Structure

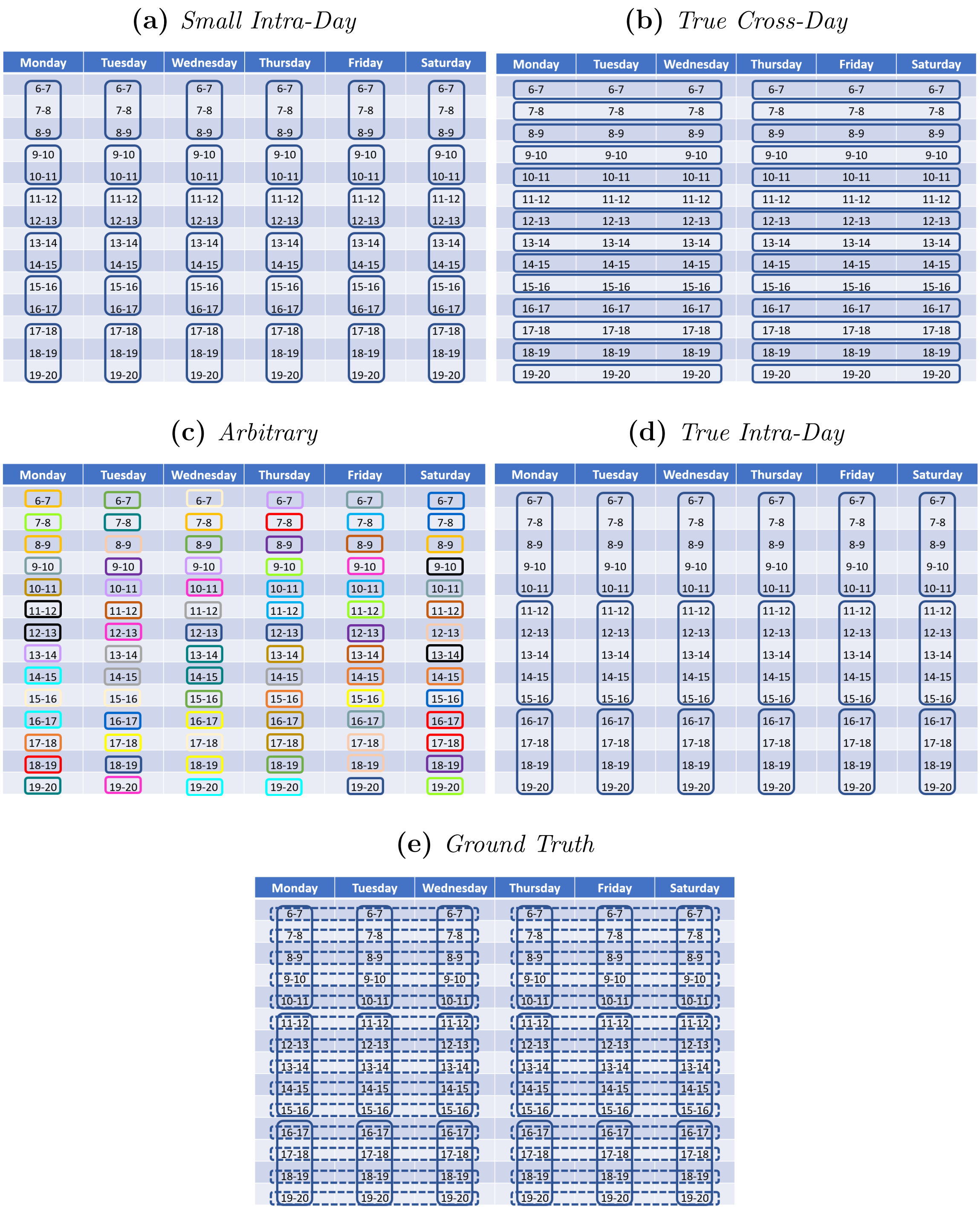


2.6.2.1. Impact of the Predefined Cluster Structure

In this subsection, we seek to investigate how the predefined cluster structure used in the estimation process of our consider-then-choose customer choice model impacts estimation quality. For this purpose, we vary the predefined cluster structure and compare the quality of the corresponding estimated choice models. In particular, we report the results from five cluster structures denoted as *Small Intra-Day*, *True Cross-Day*, *Arbitrary*, *True Intra-Day*, and *Ground Truth*. The clusters are visualized in Figure 2.2 and outlined below.

- The cluster structure *Small Intra-Day* includes time slot clusters in the early morning, i.e., from 6:00 to 9:00, in the morning, i.e., from 9:00 to 11:00, around lunch time, i.e., from 11:00 to 13:00, in the early afternoon, i.e., from 13:00 to 15:00, in the late afternoon, i.e., from 15:00 to 17:00, and in the evening, i.e., from 17:00 to 20:00, for every day.
- The cluster structure *True Cross-Day* consists of a specific time slot in the first or second half of the week, i.e., from Monday-Wednesday, or from Thursday-Sunday. These clusters are identical to the ground truth model, except that no intra-day preferences are included.
- The cluster structure *Arbitrary* does not follow any screening rule. It consists of disjoint clusters with 4 random time slots each.

Figure 2.2.: Visualization: Cluster Structures



- Cluster structure *True Intra-Day* follows the screening rule of considering time slots of one day and partitions the time slots into three blocks. In particular, it includes time slots at the beginning of the day, i.e., from 6:00 to 11:00, the middle of the day, i.e., 11:00 to 16:00, and the end of the day, i.e., from 16:00 to 20:00. These clusters are identical to the ground truth model, except that no cross-day preferences are considered.

- The cluster structure *Ground Truth* uses identical screening rules as the ground truth model and forms its clusters accordingly.

Table 2.1 provides the root mean squared error (RMSE) and the mean absolute deviation (MAD) of the estimated choice probabilities for the different cluster structures, compared to the ground truth choice probabilities.

Table 2.1.: Estimation Quality: Different Predefined Cluster Structures

Cluster Structure	RMSE	MAD	Computation Time
Small Intra-Day	0.0682 (±0.0006)	0.0395 (±0.0003)	3.82 (±0.24)
True Cross-Day	0.0528 (±0.0009)	0.0310 (±0.0005)	4.01 (±0.38)
Arbitrary	0.0447 (±0.0012)	0.0282 (±0.0005)	104.74 (±5.31)
True Intra-Day	0.0243 (±0.0016)	0.0168 (±0.0010)	1639.31 (±150.10)
Ground Truth	0.0230 (±0.0012)	0.0160 (±0.0009)	2159.31 (±120.10)

To visualize the results of the estimation procedure, we include scatter plots, as in Van Ryzin and Vulcano (2015), that depict the ground truth choice probabilities versus the probabilities from the estimated consider-then-choose model, i.e., pairs

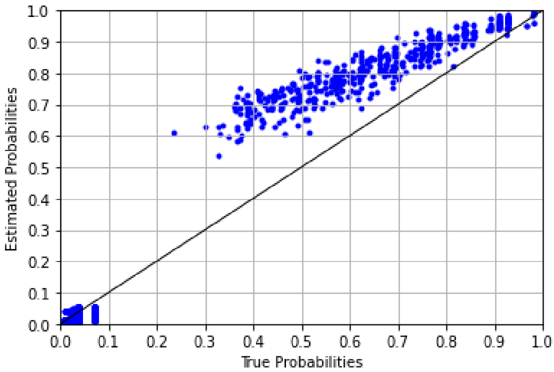
$$(\mathbb{P}_{GroundTruth}(j_t, S_t),\, \mathbb{P}_{Estimated}(j_t, S_t)),\; j_t \in S_t$$

for the test assortments. The 45° line represents perfect fit. Figure 2.3 displays the results for the considered cluster structures.

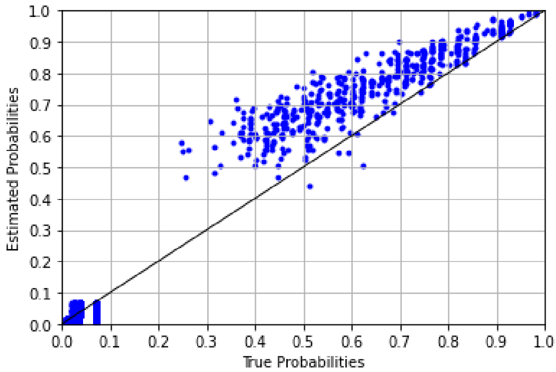
The results in Table 2.1 document how the assumed cluster structure impacts the estimation quality of the consider-then-choose choice model. We observe, unsurprisingly, that the cluster structure *Ground Truth* yields the best estimates whereas the cluster structure *Small Intra-Day* yields the poorest estimates, both in terms of RMSE and MAD. We note that the performance of the cluster structure *True Intra-Day* is insignificantly worse than the cluster structure *Ground Truth*, whereas the cluster structure

Figure 2.3.: Estimation Results: Cluster Structures

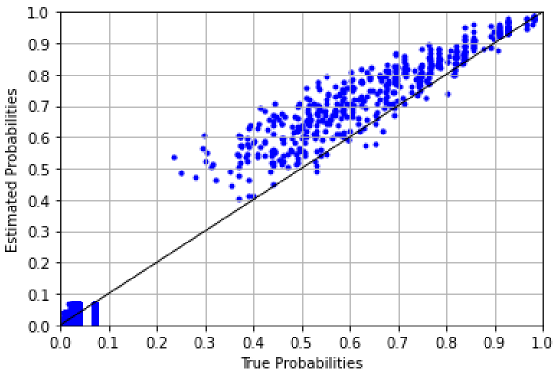
(a) *Small Intra-Day*



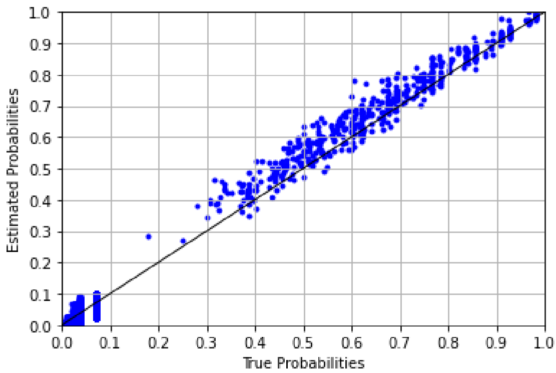
(b) *True Cross-Day*



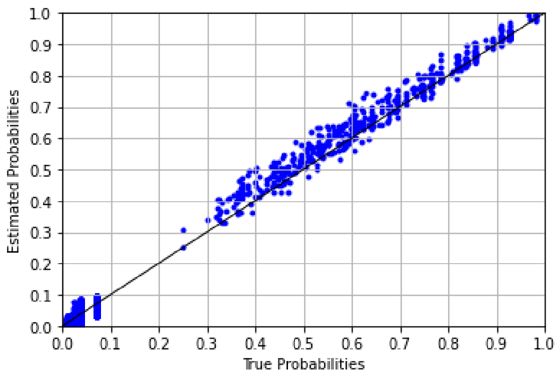
(c) *Arbitrary*



(d) *True Intra-Day*



(e) *Ground Truth*

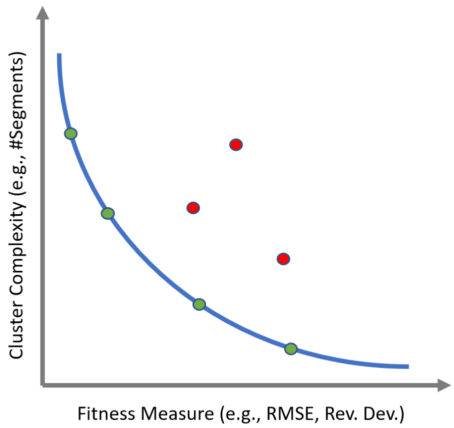


True Cross-Day provides poor results. Counter-intuitively, we note that the RMSE and MAD of the cluster structure *Arbitrary*, which does not follow any pattern for creating the clusters, are lower than those from the cluster structures *Small Intra-Day* and *True Cross-Day* (Table 2.1, Figure 2.3). This indicates that, in the given setting, the number of products considered in the clusters is particularly important, as it grants customers a higher degree of product substitutability. In Figure 2.3, we observe a bias, especially for cluster structures *Small Intra-Day* and *True Cross-Day* (Figure 2.3a, 2.3b), toward underestimating low purchase probabilities and overestimating high purchase probabilities. A potential reason is that the assumed cluster structures do not provide sufficient flexibility for product substitution. Since customers within a cluster consider only the products specified by the cluster, it is possible that none of the considered products is offered in an assortment. This lack of substitutability of products in the cluster structure leads to an underestimation of the purchase probabilities of the considered products (in the range of 0%-10%) and in turn to an overestimation of the no-purchase option (in the range of 30%-100%). In the extreme setting of an independent demand model, which can be captured in our setting by clusters considering only a single product, this bias is amplified and was investigated in Van Ryzin and Vulcano (2015). Note that the above analysis addresses only the deviations in purchase probabilities. In Section 2.6.3.1, we investigate the impact on the revenue achieved in assortment optimization.

2.6.2.2. Designing the Cluster Structure

Finding an appropriate cluster structure in which to embed the estimation process and subsequently the assortment optimization is not trivial. One way to obtain a suitable cluster structure is a two-step procedure that generates various plausible cluster structures in a first step and selects the "best" cluster structure in a second step. A good starting point for designing cluster structures is to use common screening rules (Gilbride and Allenby 2004, Jagabathula and Rusmevichientong 2017). Depending on the considered screening rules, this approach leads to several candidate cluster structures. A fitness measure (e.g., RMSE for estimation or revenue deviation for assortment optimization) can be used to evaluate these cluster structures. Complex cluster structures typically lead to better fitness measures. On the other hand, more complex cluster structures increase the difficulty of the estimation procedure (Section 2.6.2.1) and the online computation time of the optimization (Section 2.6.3.1). This trade-off between fitness measure and cluster complexity should be carefully considered (Figure 2.4).

Figure 2.4.: Efficient Frontier



2.6.2.3. Impact of the Ground Truth Choice Model

In this section, we explore the performance of our proposed estimation approach for alternative underlying customer choice behavior. To this end, we use a parametric customer choice model as the ground truth. To obtain realistic instances, we use the cross-nested logit (CNL) choice model based on parameters estimated from real booking data of a large e-grocer in the UK (Yang et al. 2016). As described in Farias et al. (2013), in the CNL model, the products (except the no-purchase option) are partitioned into B disjoint nests, denoted by $\mathcal{N}_1, \ldots, \mathcal{N}_B$, i.e., $\mathcal{N} = \bigcup_{b=1}^{B} \mathcal{N}_b$ with $\mathcal{N}_a \cap \mathcal{N}_b = \emptyset$ for $a \neq b$. In contrast to the ground truth customer choice model used hitherto, this model does not involve any clusters, but rather prescribed nests. The probability that product j is chosen from an assortment A is given by

$$\begin{aligned}\mathcal{P}(j \mid A) &= \mathcal{P}(\mathcal{N}_b \mid A) \cdot \mathcal{P}(j \mid \mathcal{N}_b,\, A)\\ &= \frac{(w(b, A))^{\rho}}{\sum_{b=1}^{B} (w(b, A))^{\rho}} \cdot \frac{w_j}{w(b, A)}\end{aligned}$$

with a scale parameter $\rho < 1$ and

$$w(b, A) = \alpha_b w_0 + \sum_{i \in (\mathcal{N}_b \cap A) \setminus \{0\}} w_i.$$

The parameters $\alpha_b \geq 0$ are the level of membership of the no-purchase option in nest b, and the following holds

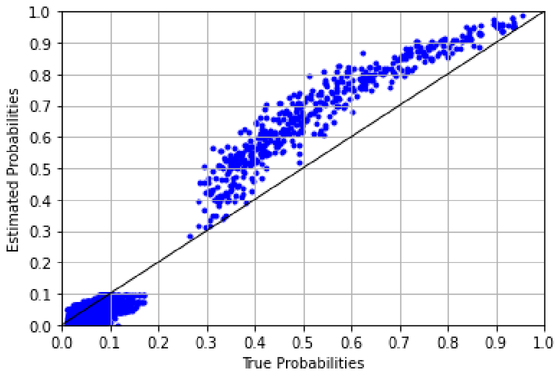
$$\sum_{b=1}^{B} \alpha_b^{\rho} = 1.$$

To mimic realistic customer choice behavior, we use the weights w_j from Yang et al. (2016). To incorporate these values into the CNL choice model, we use a scale parameter $\rho = 0.5$, 6 nests representing each delivery day of the week, and allocate the no-purchase option equally to each nest. Table 2.2 provides the RMSE and MAD of our estimation approach assuming the *Ground Truth* but using the CNL for the actual customer behavior. Figure 2.5 visualizes the findings.

Table 2.2.: Estimation Quality: CNL Ground Truth Customer Choice Model

Ground Truth	**RMSE**	**MAD**	**Computation Time**
CNL	0.0578 (±0.0007)	0.0371 (±0.0005)	2597.22 (±271.95)

Figure 2.5.: Estimation Results: CNL Ground Truth



We observe that the RMSE and MAD are relatively high, especially when comparing these figures to those obtained when using a consider-then-choose ground truth model (Section 2.6.2.1). This is not surprising, as the estimation method is tailored to extract the observed customer choice behavior for a non-parametric customer choice model. Nevertheless, although the underlying customer choice model is structurally different, the tuples $(\mathbb{P}_{GroundTruth}, \mathbb{P}_{Estimated})$ are still strongly correlated (Figure 2.5).

Insights: Estimation

In summary, we observe that our estimation method is able to identify the underlying customer choice behavior for many time slots over multiple days from historical transaction data. The estimation of the customer choice model seems to work better for cluster structures that provide a certain degree of product substitutability. For all different predefined cluster structures with larger clusters in our numerical study, the estimated consider-then-choose customer choice model captures the underlying choice behavior. In the next subsection, we investigate the resulting performance of the corresponding assortment optimization.

2.6.3. Assortment Optimization

To assess whether our proposed approach is applicable in the considered context of offering time slots, we investigate two aspects, namely the computation time (Section 2.6.3.1) and the achieved revenues (Section 2.6.3.2). We use the model formulation (2.5) - (2.8) with single-element intersection constraints (2.7) for all our numerical experiments and solve all problems to optimality.

To evaluate the runtime of our model formulation, we benchmark our assortment optimization formulation against the integer programming formulation used in Aouad et al. (2020). In that formulation, two different classes of decision variables are introduced. The first class are binary variables $y_i \in \{0, 1\}$ that decide whether a product $i \in \{1, \ldots, N\}$ is included in the assortment. Compared to the model formulation proposed in this paper, the model explicitly determines the total assortment on product level for all customer segments at once. The second class of decision variables $x_{l,i} \in \{0, 1\}$ indicate whether customer segment l chooses product i. The complete problem is formulated as follows:

$$\max \sum_{i=1}^{N} \sum_{l=1}^{L} \lambda_l \cdot r_i \cdot x_{l,i} \tag{2.12}$$

$$x_{l,i} \leq y_i \qquad \forall\, l \in \{1, \ldots, L\}, \forall\, i \in C_l \tag{2.13}$$

$$x_{l,i} + y_m \leq 1 \qquad \forall\, l \in \{1, \ldots, L\}, m \in C_l \text{ and } i \in \{x \in C_l : \sigma_l(m) < \sigma_l(i)\} \tag{2.14}$$

$$\sum_{i \in C_l} x_{l,i} \leq 1 \qquad \forall\, l \in \{1, \ldots, L\} \tag{2.15}$$

$$x_{l,i} \in \{0, 1\} \qquad \forall\, l \in \{1, \ldots, L\}, \forall\, i \in C_l \tag{2.16}$$

$$y_i \in \{0, 1\} \qquad \forall\, i \in \{1, \ldots, N\}. \tag{2.17}$$

The objective function (2.12) maximizes the weighted sum of revenues received from the customer segments. The first class of constraints (2.13) ensures that a customer segment l can only choose a product i if the product is included in the assortment. The second class of constraints (2.14) guarantees that a customer segment always chooses the most preferred product from the offered assortment. Inequalities (2.15) ensure that each customer segment chooses at most one product. Therefore, the no-purchase option is also implicitly included in this model.

2.6.3.1. Computational Efficiency

Random Customer Choice Model with Cluster Structure

In this subsection, we optimize the assortment under randomly generated consider-then-choose customer choice models with an underlying cluster structure and compare the computation times of our model formulation with those of the benchmark formulation. We also investigate how the results depend on the underlying cluster structure and on the size and the number of clusters. To generate these choice models, we assume that each cluster contains multiple customer segments and that the preference lists of the included customer segments are given as all permutations of the products. As in Aouad et al. (2020), we assume equal weights for the different customer segments.

Cluster Size

First, we evaluate the impact of the cluster sizes on the computation time of both optimization formulations. We consider a simple underlying cluster structure with only two customer clusters of similar size. Since the difference in computation times becomes evident even for a small number of products, the clusters consider 5, 6, 7, and 8 products, respectively, and have an intersection of three products. Table 2.3 shows the computation times of our optimization approach and the benchmark.

Table 2.3.: Computation Times: Different Cluster Sizes

Cluster Size	Preprocessing	Runtime	Benchmark
5	0.06 ($\pm$0.01)	0.04 ($\pm$0.01)	1.49 ($\pm$0.16)
6	0.84 ($\pm$0.02)	0.04 ($\pm$0.01)	10.42 ($\pm$2.79)
7	14.47 ($\pm$0.77)	0.04 ($\pm$0.01)	126.04 ($\pm$3.83)
8	252.08 ($\pm$7.61)	0.04 ($\pm$0.01)	$> 10^3$

We observe that the total computation times of our proposed optimization formulation are significantly lower than those for the benchmark formulation, for all considered cluster sizes. For both optimization formulations, the computation times increase with increasing cluster sizes. The same applies to the difference in computation time. It is notable that the online runtime of our model formulation remains constant at 0.04 seconds. In all settings, most of the computation time of our proposed optimization approach is not required for the final, online optimization but for preprocessing the data, i.e., evaluating the cluster revenues. In the application context we consider, where the assortment problem has to be solved quickly, this is a considerable advantage.

Number of Clusters

Second, we examine the impact of the number of clusters. To this end, we consider an underlying cluster structure with up to eight customer clusters, each considering 6 products, and having an overlap of two products each. Table 2.4 reveals the computation times of both approaches.

Again, it is apparent that the computation times of our approach are significantly shorter than those of the benchmark for all considered numbers of clusters. As observed previously, most of the computation time is used for preprocessing the data. The difference in computation times increases as the number of clusters increases, implying that the computation time advantage of our approach compared to the benchmark increases as the number of clusters increases.

Table 2.4.: Computation Times: Different Number of Clusters

Number Clusters	Preprocessing	Runtime	Benchmark
2	0.92 ($\pm$0.03)	0.04 ($\pm$0.01)	10.97 ($\pm$1.65)
4	1.68 ($\pm$0.06)	0.05 ($\pm$0.01)	21.32 ($\pm$2.63)
8	2.99 ($\pm$0.06)	0.08 ($\pm$0.01)	34.93 ($\pm$2.61)

Customer Choice Model with Cluster Structure from Estimation Procedure
In addition to randomly generated consider-then-choose customer choice models, we also apply our optimization approach to our estimated choice models from Section 2.6.2.1. We investigate the computation times of both optimization formulations for realistic problem sizes of an e-grocer. Table 2.5 displays the computation times for different predefined cluster structures.

Table 2.5.: Computation Times: Different Cluster Structures used in the Estimation

Cluster Structure	Preprocessing	Runtime	Benchmark
Small Intra-Day	0.01 ($\pm$0.01)	0.07 ($\pm$0.01)	0.22 ($\pm$0.01)
True Cross-Day	0.01 ($\pm$0.01)	0.06 ($\pm$0.01)	0.27 ($\pm$0.01)
Arbitrary	0.05 ($\pm$0.01)	0.05 ($\pm$0.01)	1.09 ($\pm$0.03)
True Intra-Day	0.37 ($\pm$0.01)	0.07 ($\pm$0.01)	7.05 ($\pm$0.05)
Ground Truth	0.36 ($\pm$0.01)	0.20 ($\pm$0.01)	7.29 ($\pm$0.10)

The figures show that the computation times of our approach are consistently lower than those of the benchmark for all predefined cluster structures. Regarding the randomly generated customer choice models, the differences in computation times vary, depending on the underlying cluster structure (different number of clusters, different cluster size). For the three cluster structures *Small Intra-Day*, *True Cross-Day*, and *Arbitrary* with the lower estimation quality, the difference in computation times is not as high as for the cluster structures *True Intra-Day* and *Ground Truth* with higher estimation quality. In general, however, the differences between the computation times are not as extreme as for randomly generated data. The reason is that not all customer segments are activated (and obtain a positive weight) due to the estimation procedure, whereas this was the case for randomly generated data. Nevertheless, the computation times of the benchmark formulation are too high for practical implementation, whereas an online time slot optimization would be possible with our model formulation.

2.6.3.2. Revenue Impact of the Cluster Structure

In this subsection, we examine the impact of assuming different customer cluster structures in the estimation procedure on the achieved revenues in the assortment optimization. We again use the estimated customer choice models from Section 2.6.2.1. We simulate arriving customers and different products with different revenues and determine the optimal assortment for the estimated customer choice model. We then use the ground truth model to determine the true optimal assortment and the true expected revenues obtained from the calculated assortments. Table 2.6 displays the expected revenue deviations in % between the assortments obtained from the estimated customer choice models and the optimal assortment using the ground truth model.

Table 2.6.: Impact on Revenue: Different Cluster Structures used in the Estimation

Cluster Structure	**Revenue Deviation in %**
Small Intra-Day	31.01 (±1.01)
True Cross-Day	28.72 (±1.57)
Arbitrary	29.43 (±1.58)
True Intra-Day	0.36 (±0.05)
Ground Truth	0.75 (±0.24)

The figures indicate strong performance differences between the assumed cluster structures. Essentially, we can divide the clusters into two main performance groups. The greatest revenue deviations are in the range of approximately 30% and can be observed if the cluster structure *Small Intra-Day*, *True Cross-Day*, or *Arbitrary* is used in the estimation procedure. It is not surprising that these cluster structures yield worse results because we observed a lower customer choice estimation quality (Table 2.1, Figure 2.3), which naturally leads to a larger deviation in revenue. The figures in Table 2.6 confirm that the purchase probabilities of crucial products were estimated incorrectly, leading to a poor assortment decision. In contrast, the estimated consider-then-choose customer choice models using the predefined cluster structures *True Intra-Day* and *Ground Truth*, which exhibit high estimation quality (Table 2.1, Figure 2.3), yield the lowest revenue deviations of less than 1% (Table 2.6). This corresponds to the identification of the true optimal assortment in almost all simulations and demonstrates that the purchase probabilities were estimated sufficiently well to determine the optimal assortment decision.

Insights: Optimization

To summarize, our numerical study indicates that our optimization formulation performs favorably for instances with an underlying customer cluster structure. In all experiments, the computation times for our proposed formulation are significantly shorter than those of the benchmark. The computation time advantage increases with increasing cluster size and an increasing number of clusters. Most important, only a fraction of the total computation time is spent on the online optimization and most computational effort can be performed as an offline pre-processing step. This is especially beneficial for our application context. Using this method, an e-grocer can solve the online time slot offering problem with many time slots over multiple days in real-time. We also confirmed that expected revenues from the assortment optimization depend on the assumed cluster structure. A more accurate estimation of the customer choice behavior generally results in higher revenues. Nevertheless, there may be cluster structures other than the true underlying cluster structure that perform well and lead to the optimal assortment in almost all simulations.

2.7. Conclusion

To conclude, we investigate an e-grocer's online assortment problem of offering time slots to an arriving customer from a customer choice perspective. We propose a tractable assortment optimization approach for an underlying consider-then-choose customer choice model that is based on aggregated product level and leverages an underlying customer cluster structure. To obtain this customer choice model, we propose an estimation procedure for determining a consider-then-choose customer choice model ensuring a predefined customer cluster structure based on historical transaction data. The estimation procedure and the assortment optimization approach are investigated in a numerical study. We conclude that the estimation of the consider-then-choose choice model can identify the underlying customer behavior from historical transaction data. In particular, we observe that the estimation works well if the predefined cluster structure allows a certain degree of product substitutability. Our optimization formulation leads to significantly shorter total computation times and, most important, online runtimes, which is especially favorable in our application context.

This paper opens new avenues for future research. In our studies, we tested both the estimation procedure and the assortment optimization approach on simulated realistic data. Applying our approaches to real transaction data from e-groceries would be interesting to validate our findings. Similarly, it would be interesting to test the different predefined customer cluster structures with real data and investigate whether the degree of substitutability of our predefined cluster structures is sufficient. Future research could address the screening/preselection rules used to define and optimize the assumed cluster structure, for example, by integrating them into the estimation process. This would significantly complicate the estimation procedure but may be worth exploring. In our study, we focused on the time slot offering problem from a customer choice perspective and assumed all routing constraints to be given. Future research could integrate these two problem components by combining our online time slot optimization approach for an underlying consider-then-choose choice model with a sophisticated routing model.

Chapter 3.

Feeding the Nation - Dynamic Customer Contacting for E-Fulfillment in Times of Crisis[1]

with Moritz Fleischmann and Arne Strauss

[1]The research presented in this chapter is based on a paper entitled "Feeding the Nation - Dynamic Customer Contacting for E-Fulfillment in Times of Crisis", coauthered with Moritz Fleischmann and Arne Strauss (Schwamberger et al. 2022b).

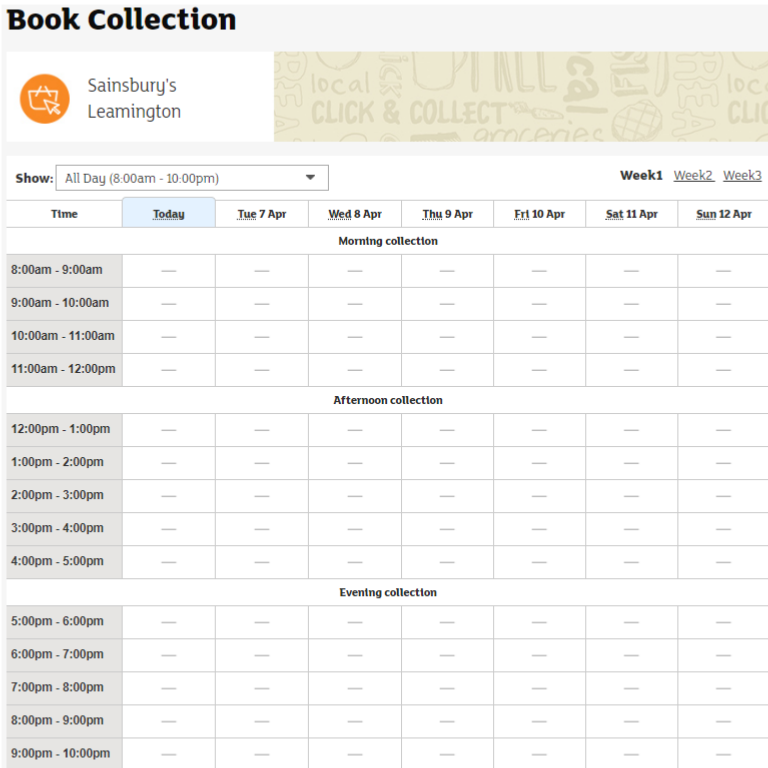
Abstract

The outbreak of the COVID-19 pandemic led demand for online grocery orders for both C&C and AHD to outstrip delivery capacity by a wide margin. In the UK, the booking systems of some e-retailers could not handle the flood of incoming requests, forcing the retailers to proactively reach out to certain priority customer segments with the aim of serving as many high-priority customers as possible. To determine when to contact each customer segment in this extraordinary demand environment, we investigate the new demand management concept of proactively contacting customers. We first develop a decision policy for the C&C fulfillment method to address the problem of when to contact customers. We then extend this approach to the AHD setting. To cope with increased problem complexity, we propose a three-step procedure to solve this problem. First, we subdivide the delivery area into smaller subareas; second, we select the most promising subareas; and third, we determine which customers to contact within the chosen subareas. To gain managerial insights and to show the practical benefits of our approaches, we apply both approaches to realistic data from the London area. Our results show that proactive customer contacting allows for a tailored allocation of scarce capacity in e-groceries. We conclude with a discussion on how these contacting techniques can be valuable in post-crisis times.

3.1. Introduction

With the outbreak of the COVID-19 pandemic, life has changed rapidly in most places worldwide. With rising incidence rates and an increased infection risk, many governments have encouraged residents to limit their social contacts to a minimum. Companies have reacted accordingly, and many employees have been given the opportunity to work from home. As a result of this shutdown of public life by governments and corporations, many activities have shifted toward the digital world. Due to this shift, online shopping has increased dramatically. This holds true especially for the proportion of groceries purchased online, where sales grew by 75% in 2020 (Mintel 2021) and spiked even much more dramatically (McKinsey 2020). Consequently, the demand for online grocery orders far outstripped capacity in many markets worldwide, resulting in widely sold-out time slots (CNA 2020, Telegraph 2020, Vox 2020), both for C&C and AHD services (Figure 3.1, Figure 3.2).

Figure 3.1.: Sainsbury's Time Slot Offering During COVID-19 Pandemic



Book Collection

Sainsbury's Leamington

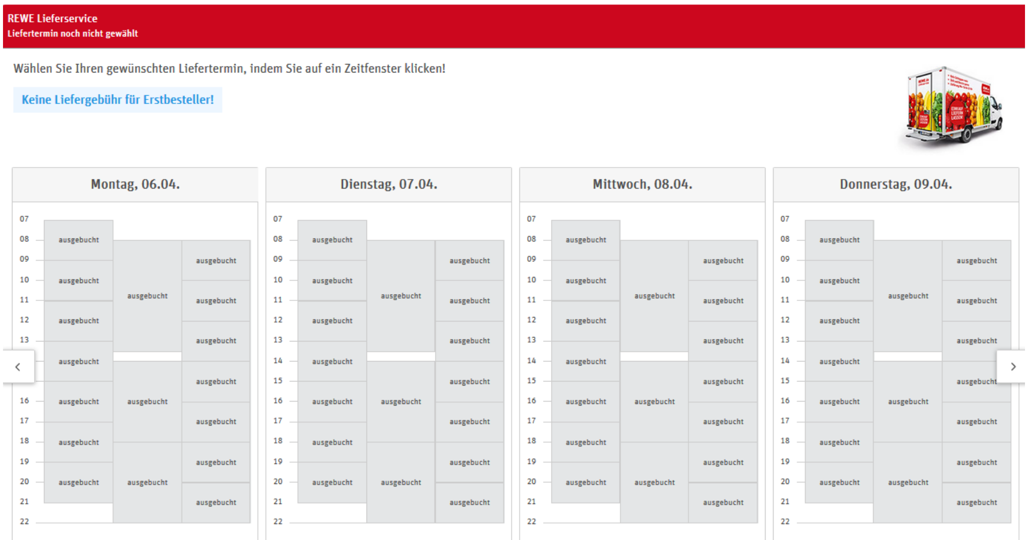
local CLICK & COLLECT groceries

Show: All Day (8:00am - 10:00pm)

Week1 Week2 Week3

Time	Today	Tue 7 Apr	Wed 8 Apr	Thu 9 Apr	Fri 10 Apr	Sat 11 Apr	Sun 12 Apr
Morning collection							
8:00am - 9:00am	—	—	—	—	—	—	—
9:00am - 10:00am	—	—	—	—	—	—	—
10:00am - 11:00am	—	—	—	—	—	—	—
11:00am - 12:00pm	—	—	—	—	—	—	—
Afternoon collection							
12:00pm - 1:00pm	—	—	—	—	—	—	—
1:00pm - 2:00pm	—	—	—	—	—	—	—
2:00pm - 3:00pm	—	—	—	—	—	—	—
3:00pm - 4:00pm	—	—	—	—	—	—	—
4:00pm - 5:00pm	—	—	—	—	—	—	—
Evening collection							
5:00pm - 6:00pm	—	—	—	—	—	—	—
6:00pm - 7:00pm	—	—	—	—	—	—	—
7:00pm - 8:00pm	—	—	—	—	—	—	—
8:00pm - 9:00pm	—	—	—	—	—	—	—
9:00pm - 10:00pm	—	—	—	—	—	—	—

Figure 3.2.: Rewe's Time Slot Offering During COVID-19 Pandemic



The extreme competition for online grocery ordering raises the question of which customers to prioritize, given the scarce capacity. During the COVID-19 pandemic, many retailers have chosen to not simply prioritize customers based on short-term profitability. They instead emphasize their role as essential service providers that "work to feed the nation" (as Sainsbury's claims on its website). In this vein, they seek to assure access for particularly vulnerable customer groups. In the UK, for example, retailers have been collaborating with the government to define who qualifies for priority access as elderly, disabled & vulnerable (EDV) (Department of Finance, UK 2020). Some retailers likewise treated medical staff as a priority group (Rewe 2020). In addition, retailers commonly prioritized valuable existing customers (e.g., those who hold a premium pass for online grocery orders).

One way to implement priority access for selected customer groups is to simply deny orders from non-priority customers. However, this is inefficient as it may still overload the booking system with many failed ordering attempts and cause frustration to customers due to excessive waiting times during the ordering process. Therefore, retailers have been considering an alternative approach, namely to proactively reach out to certain priority customer segments (Yorkshire Post 2020). This approach is also explored by our industry partner.

The concept of proactively contacting customers raises several questions: Which customers should be contacted and when? How does the customer contacting affect relevant performance metrics? How should the contacting approach differ, dependent on the design of the fulfillment process? How does customer contacting perform relative to traditional reactive order intake? These are the questions that we address in this paper. Specifically, we explore the new contacting concept, propose solutions for C&C and AHD fulfillment channels, and evaluate their performance and compare them with the status quo.

Our main contributions to the literature include the following: (1) the development of a dynamic programming-based solution approach for C&C, (2) documenting the practical benefit of proactively contacting customers with realistic data, and (3) the development of a corresponding procedure for the more challenging AHD channel. This procedure consists of three steps: first, subdividing the delivery area into smaller subareas; second, selecting the most promising subareas; and third, determining which customers to contact in the selected subareas. As in the C&C setting, we conduct a numerical study on realistic data in the London area to gain managerial insights and evaluate the merit of proactive customer contacting in this setting. Furthermore, we discuss how these contacting techniques can be valuable in post-crisis times.

The paper is organized as follows: in Section 3.2, we review and discuss the related literature. In Section 3.3, we discuss the underlying problem setting and Section 3.4 presents the solution approaches for proactively contacting customers in C&C and AHD. In Section 3.5, we present the results of our numerical studies with realistic data from the London area. In Section 3.6, we provide decision support as to when an e-grocer should switch to proactive customer contacting and conclude with a discussion how the proposed contacting techniques can be valuable in post-crisis times. Section 3.7 summarizes our results.

3.2. Related Literature

The work presented in this paper interfaces with multiple research streams. In addition to the recent stream on implications of the current COVID-19 pandemic, these include demand management for e-grocery services, customer relationship management, revenue management, routing literature on the selective traveling salesman problem (STSP), the team orienteering problem (TOP), and inventory management with random yields. We briefly review each of these streams and position our work relative to it.

The COVID-19 pandemic has brought substantial challenges for supply chain management. Many new problems for which an established scientific understanding was lacking had to be solved quickly. One and a half years later, researchers in various fields have been addressing crisis-related problems and have made contributions to solving them in more systematic ways (see, e.g., Pavlik et al. (2021) and Choi (2020)). However, to the best of our knowledge, no work has yet addressed the impact of the COVID-19 crisis on the distribution processes of e-groceries. With this paper, we contribute to this relevant field.

In recent years, demand management has gained increasing importance in last-mile distribution systems, e.g., Agatz et al. (2011), Yang et al. (2016), and Yildiz and Savelsbergh (2020). Agatz et al. (2013) identified time slot allocation and time slot pricing as the two main levers of demand management in the context of e-grocers offering AHD and C&C services. These decisions are either made in a dynamic or a static fashion. A good overview of recent work in the field of demand management for e-grocers with an AHD service can be found in the literature section of Mackert (2019), resulting in the four cases of differentiated slotting (Agatz et al. 2011, Cleophas and Ehmke 2014), differentiated pricing (Klein et al. 2017), dynamic slotting (Ehmke and Campbell 2014, Mackert 2019) and dynamic pricing (Yang and Strauss 2017, Klein et al. 2018). However, none of these studies included the option of proactively contacting customers, which is the key issue addressed in this paper. To the best of our knowledge, proactive customer contacting has only been investigated in Keskin et al. (2023) and Yildiz and Savelsbergh (2020). Keskin et al. (2023) considered proactive customer contacting in the context of waste collection and investigated the trade-off between more efficient routes and smaller collection volumes resulting from the increased frequency of visits generated by contacting customers. We, in contrast, focus on the contacting problem in the context of e-groceries and investigate how to incorporate the prioritization of certain customers in a scarce capacity situation. Yildiz and Savelsbergh (2020) considered customer contacting in the context of e-groceries, but they assumed that the time slot pricing decision is the main lever of demand management and examined the impact of contacting customers by offering discounts in exchange for delivery flexibility. Although they suggested actively contacting customers, their contacting setting is quite different from ours, as customers were only contacted after they placed an order so as to move their order to a different time slot. In our problem setting, the e-grocer starts with a base of potential customers that can be contacted, thereby endogenizing the customer arrival process. In addition, we focus on the time slot offering rather than the pricing decision. For a more detailed overview of managerial planning tasks in this context, we recommend Agatz et al. (2008b), who reviewed the e-fulfillment process and discussed corresponding quantitative models. Recent advantages of integrating demand management and vehicle routing can be found in the review of Fleckenstein et al. (2023). However, the proactive contacting method that we propose in this paper has not yet been studied in this stream of literature.

Since the e-grocer decides which customers have the opportunity to place an order in the first place by proactively contacting customers, our research interfaces with customer relationship management (Landrigan 2005). In particular, concepts such as customer lifetime value should be considered in the contacting and prioritization decisions of the e-retailers (Berger and Nasr 1998). In this regard, subscriptions and delivery passes could be used to define the customer's priority status. In addition, our study is related to research on direct marketing and its associated marketing issues, as the e-grocer directly addresses customers and offers a certain assortment of time slots in our proposed approach. However, we do not focus on the customer relationship and marketing issues associated with this contacting and promoting process but rather address the optimization issues underlying this process. For an overview of decision models for customer relationship management and quantitative models for direct marketing, we refer to Reinartz and Venkatesan (2008) and Bose and Chen (2009).

Our work is also related to the domain of revenue management. In particular, we encounter a version of the classic overbooking problem under the assumption of stochastic customer responses such that we need to decide how many customers to contact beyond our anticipated delivery capacity in anticipation of some customers not responding (yet we experience an ill-will penalty if we reject customers who respond but cannot be served due to insufficient capacity). In contrast to the classic overbooking problem, we do not know the exact delivery capacity under attended home delivery since this depends on the location and order volumes of all responding customers. Klein et al. (2020) discussed the general overbooking problem and presented revenue management techniques in online grocery retailing.

As our study incorporates routing considerations in the AHD setting and it needs to be determined which customers to optimally contact in a deterministic customer arrival setting, it intersects with the literature on the STSP and the TOP. In particular, we modify the TOP formulation of Tang and Miller-Hooks (2005) to obtain the optimal set of customers to contact and serve in the deterministic arrival setting. For a more detailed overview of orienteering problems, we refer to Vansteenwegen et al. (2011). Slightly less obviously, our study also shows some aspects of inventory control models with random yields. This analogy stems from the fact that it is uncertain whether a customer contacted by an e-grocer actually accepts the invitation and places an order. For detailed reviews on the random yield literature, we refer to Grosfeld-Nir and Gerchak (2004) and Yano and Lee (1995).

In summary, the analysis presented in this paper draws inspiration from the different research streams outlined above. However, no previous research has addressed the relevant planning problem of an e-grocer seeking to manage demand in a surge situation by proactively contacting potential customers. Introducing and analyzing this novel planning problem is how our paper contributes to the literature.

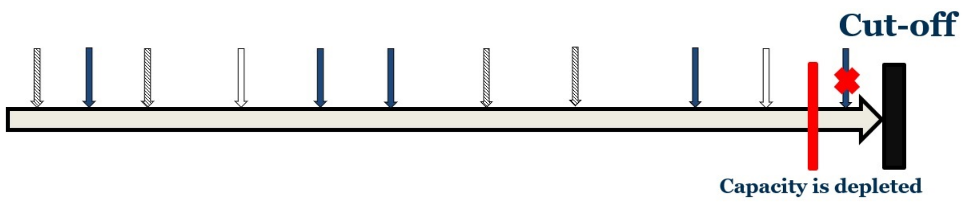
3.3. Problem Setting

In this section, we formalize the problem setting for C&C and AHD and highlight the main differences between both fulfillment methods.

3.3.1. Click&Collect

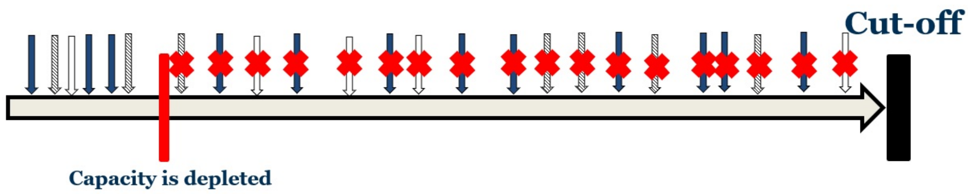
We first focus on an e-grocer offering a C&C service. For this service, the system capacity is largely determined by store operations. In non-crisis times, there is no need to prioritize customers in the booking process, as capacity is sufficient to serve most customers. Thus, customers arrive randomly throughout the booking horizon and are simply served according to the first come first served (FCFS) principle (Figure 3.3).

Figure 3.3.: C&C Booking Process in Normal Demand Situations



During the COVID-19 crisis, demand has far exceeded capacity. While e-retailers can increase handling capacity to some extent, they are unable to cover the enormous demand surge. This has resulted in many dissatisfied customers who are not able to select a time slot at all and thus cannot be served (Figure 3.4). From an e-retailer's point of view, the FCFS method is no longer satisfactory since in such an extreme demand situation, customer-specific priority affiliations cannot be incorporated into the booking process despite their increasing importance.

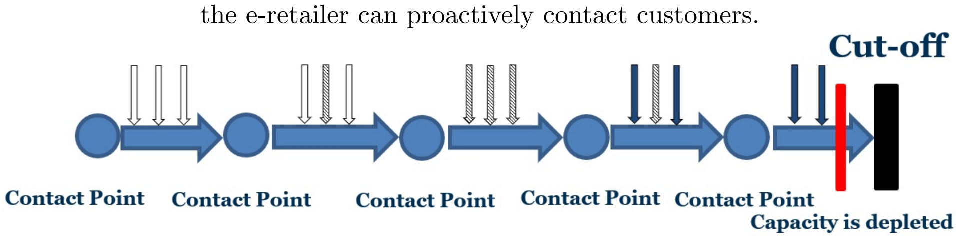
Figure 3.4.: C&C Booking Process in Extreme Demand Situations



By proactively contacting customers, the e-retailer can manage the customer arrival process in extraordinary demand situations to incorporate customer segment affiliations (Figure 3.5). Each customer c is associated with a segment-specific weight α_c that denotes the importance of serving customer c.

Figure 3.5.: Contacting Procedure

Our proposed contacting procedure divides the time horizon into several contacting periods in which the e-retailer can proactively contact customers.



While in our numerical study the weights reflect the vulnerability of certain customer segments, akin to the current approach of several UK retailers, the definition is generic and allows for other choices, such as profit-based priorities. In our proposed contacting approach, customers can be contacted at a predefined contacting point t in one of the T contacting points that cover the booking horizon. After a customer has been contacted, an order can be placed and one of the still available time slots selected until the next contacting point, which becomes the communicated deadline for placing an order. We denote the response probability by $\lambda \in [0, 1]$ and assume that this probability is identical and independent across all customer groups. Consequently, we obtain a binomial process for the number of arriving customers in which λ denotes the response probability (of a contacted customer arriving). Since there is an ongoing crisis situation, it is reasonable to assume that customers will accept all offered time slots. This should be considered a mild assumption in the event that stay-at-home orders are issued by governments, as happened during the coronavirus pandemic. If a customer does not place an order and does not select one of the time slots by which time the next contact is made, the opportunity to book a time slot expires. The sequence in which the contacted customers arrive and place their orders between the contacting points is random. If customer contacts exceed capacity, the arriving customers are served using the FCFS principle until capacity is depleted. At the next contacting point, it becomes apparent how many of the contacted customers have arrived and actually placed an order. The e-grocer then has the possibility to contact further customers if the cut-off has not yet been reached. If too many of the contacted customers wish to place an order, a penalty will be assessed for each arriving customer who cannot book a time slot, to limit the number of contacted, unserved customers. The penalty for each rejected customer c is denoted by β_c. To reflect that rejecting customers with a high priority status, such as elderly and most vulnerable customers, is more severe since these customers are exposed to a much higher risk and rely on being able to place an order, we assume that the penalty term for a customer depends on the associated weight of the customer and can be calculated as $\beta_c = \beta \cdot \alpha_c$ with a penalty factor $\beta > 1$. We further assume that a customer is rejected upon arrival if and only if capacity is depleted. The overall objective of the contacting approach is to maximize the sum of the weights of the customers served less the total penalties. Let $\mathcal{C}_t$ denote the set of uncontacted customers and $K_t \in \mathbb{N}_0$ the remaining free capacity at the beginning of period t. We can then formulate the problem using dynamic programming (DP), using $\mathcal{C}_t$ and K_t as state variables. To this end, let $V_t(\mathcal{C}_t, K_t)$ denote the maximum expected objective to go, given the remaining set of customers $\mathcal{C}_t \neq \emptyset$ and remaining capacity $K_t > 0$ at the beginning of period t. Using this notation, we obtain for $C_t \neq \emptyset$ the Bellman equation

$$\begin{aligned} V_t(\mathcal{C}_t, K_t) = \max_{C_t \subset \mathcal{C}_t} \sum_{n=0}^{K_t} P(X_t = n) & \left[\mathbb{E} \left(\sum_{\substack{i \in B, \\ B \subset C_t, |B| = n}} \alpha_i \right) + V_{t-1}(\mathcal{C}_t \setminus C_t, (K_t - n)) \right] \\ + \sum_{n=K_t+1}^{|C_t|} P(X_t = n) & \left[\mathbb{E} \left(\sum_{\substack{i \in B, \\ B \subset C_t, |B| = K_t}} \alpha_i \right) - \mathbb{E} \left(\sum_{\substack{i \in B, \\ B \subset C_t, |B| = n - K_t}} \beta \alpha_i \right) \right], \end{aligned}$$

which can be transformed to

$$\begin{aligned}
V_t(\mathcal{C}_t, K_t) &= \max_{C_t \subset \mathcal{C}_t} \sum_{n=0}^{K_t} P(X_t = n) \left[\mathbb{E} \left(\sum_{\substack{i \in B, \\ B \subset C_t, |B|=n}} \alpha_i \right) + V_{t-1}(\mathcal{C}_t \setminus C_t, (K_t - n)) \right] \\
&\quad + \sum_{n=K_t+1}^{|C_t|} P(X_t = n) \left[\mathbb{E} \left(\sum_{\substack{i \in B, \\ B \subset C_t, |B|=K_t}} \alpha_i \right) - \mathbb{E} \left(\sum_{\substack{i \in B, \\ B \subset C_t, |B|=n-K_t}} \beta \alpha_i \right) \right] \\
&= \max_{C_t \subset \mathcal{C}_t} \sum_{n=0}^{K_t} P(X_t = n) \left[n \cdot \overline{\alpha} + V_{t-1}(\mathcal{C}_t \setminus C_t, (K_t - n)) \right] \\
&\quad + \sum_{n=K_t+1}^{|C_t|} P(X_t = n)(K_t \cdot \overline{\alpha} - \beta \cdot \overline{\alpha} \cdot (n - K_t)) \\
&= \max_{C_t \subset \mathcal{C}_t} \sum_{n=0}^{|C_t|} P(X_t = n) [\overline{\alpha} \left(\min\{n, K_t\} - \beta \cdot (n - K_t)^+ \right) \\
&\qquad\qquad + V_{t-1}(\mathcal{C}_t \setminus C_t, (K_t - n)^+)] \\
&= \max_{C_t \subset \mathcal{C}_t} \overline{\alpha} \left(\lambda \cdot |C_t| - (1 + \beta) \sum_{n=K_t+1}^{|C_t|} P(X_t = n)(n - K_t) \right) \\
&\quad + \sum_{n=0}^{K_t - 1} P(X_t = n) V_{t-1}(\mathcal{C}_t \setminus C_t, (K_t - n))
\end{aligned}$$

where $X_t \sim B(|C_t|, \lambda)$ denotes the random number of customers that respond to the invitation in period t and $\overline{\alpha} = \frac{\sum_{c \in C_t} \alpha_c}{|C_t|}$ represents the average objective weight of the set of customers C_t. Note that it holds $\mathbb{E}\left(\sum_{i \in B, B \subset C_t, |B|=n} \alpha_i\right) = n \cdot \overline{\alpha}$ since all customers have the same arrival probability. If no customer is contacted, i.e., $C_t = \emptyset$, the process simply moves to the next contacting period without any change in the state, i.e., $V_t(\mathcal{C}_t, K_t) = V_{t-1}(\mathcal{C}_t, K_t)$. The boundary conditions are $V_0(\mathcal{C}, K) = 0$ for all sets of remaining customers $\mathcal{C}$ and all remaining capacities K. Likewise, if capacity is depleted, we have $V_t(\mathcal{C}_t, 0) = 0$ for all sets of remaining customers $\mathcal{C}_t$ and, if there are no further customers to contact, $V_t(\emptyset, K_t) = 0$ for all remaining capacities K_t and for all periods t, respectively.

3.3.2. Attended Home Delivery

In practice, we observe that the majority of online orders are not fulfilled via the C&C but rather the AHD mode of service (L.E.K. 2019). Compared to the C&C fulfillment method, the retailer delivers the products to the customer's home in the AHD case, in addition to picking and preparing all products in the shopping basket. In that regard, the AHD fulfillment method can be interpreted as an extension of the C&C method, and the general booking process from a customer's perspective remains the same. Instead of selecting a pick-up station in the first step of the booking process, the customer's address is entered. Similar to the C&C setting, we propose a new demand management approach consisting of proactively contacting customers (Figure 3.5) in the AHD setting.

Although the contacting approach in the AHD setting is very similar to the C&C setting, there is a key difference between these two fulfillment methods. Under the C&C fulfillment method, only a finite physical capacity K needs to be considered. In the AHD setting, in addition to the finite physical capacity imposed by handling operations and vehicle capacities, there is also the constraint of the maximum vehicle tour length, resulting from the shifts of the vehicle, which cannot be exceeded. This additional restriction interconnects the independent customer orders and thus significantly increases the problem's complexity. Indeed, checking whether an arriving customer can be served is already a complex routing problem. To be able to solve the problem for realistic problem instances, preliminary routes for the already accepted customer orders and the nearest insertion heuristic are used to check whether adding an arriving customer is possible. Due to this additional routing component, the case may arise that an arriving customer must be rejected whereas a future arriving customer can still be served. This is not possible in the C&C setting since the resource consumption is identical across all customers.

Before diving deeper into the details of how to solve the contacting problem, it is necessary to define the general problem context of the AHD setting. We assume that the e-retailer operates $\mathcal{V}$ vehicles and uses one depot from which all customers are served. All vehicle tours start and end at this depot. As is common in the routing literature, we assume that the travel times are proportional to the travel distances, i.e., the delivery vehicles travel with a constant speed. Furthermore, we assume a fixed dwell time τ per customer. Due to the predefined shifts of the vehicles and the working time restrictions of the employees, the tour length (consisting of travel and dwell times) is also limited. The maximum tour length is denoted by $MaxLength$. Since the vehicles can only transport a certain number of crates and each order is assumed to require the same number of crates, the number of customers served on a vehicle tour is also limited. The maximum number of customers that a vehicle can serve on one tour is denoted by $MaxCap$.

3.4. Contacting Procedure

In this section, we introduce and discuss the contacting procedures for C&C and AHD. The focus is on how to handle the complexities of the AHD setting.

3.4.1. Click&Collect

To further our understanding of the C&C contacting problem, we first consider the optimal contacting policy for two special cases, namely (i) deterministic customer arrivals and (ii) a very high number of contacting points. Subsequently, we address the more realistic case of a small number of contacting points with stochastic customer arrivals.

First, consider the case in which each contacted customer places an order with certainty, i.e., $\lambda = 1$. One can view this case as an acute crisis setting during which customers are desperate to order. In this setting, the e-grocer's optimal contacting policy is very simple. Due to the deterministic customer response, the e-grocer can fill the remaining capacity exactly without overbooking. In addition, all chosen customers can be contacted simultaneously, i.e., a single contacting period suffices. In conclusion, given a deterministic customer response, it is optimal for the e-grocer to rank the customers by α_c and to immediately contact the K_t highest ranked customers.

If a customer's response to being contacted is stochastic, i.e., $\lambda < 1$, the number of available contacting periods becomes relevant. Each additional contacting opportunity allows the e-grocer to react to the observed response from the preceding contacting round. If the number of contacting periods is large (specifically if $T > \|\mathcal{C}_t\|$) the contacting problem is again simple. In fact, in this setting, it is optimal for the e-grocer to apply the optimal policy from the above deterministic case repeatedly until the capacity is depleted or all customers have been contacted. This policy again avoids overbooking.

What makes the C&C contacting problem nontrivial is the combination of a stochastic customer response and a small number of contacting periods, yet these are the conditions that hold in practice. While the probability that a contacted customer will place an order is likely to be high in a crisis, some uncertainty remains: the customer may be unavailable, may not receive the invitation, or may have already successfully used an alternative shopping opportunity. At the same time, customers cannot be contacted arbitrarily early, and need some time to respond. This limits the number of contacting periods available to the e-grocer. Under these circumstances, it is harder for the e-grocer to match demand to the available capacity. The above DP formulation captures the underlying trade-offs. For practical use in a stressful crisis situation, it is important to solve the contacting problem quickly. We therefore use the insights gained from the previously discussed special cases to simplify the problem. Specifically, we assume that the customers are contacted in a weight-descending sequence. This drastically reduces both the state and the action spaces of the DP: instead of the *set* of remaining customers $\mathcal{C}_t$, it is sufficient to track the *number* of remaining customers $|\mathcal{C}_t|$. Similarly, instead of deciding on the *set* of customers C_t to contact in period t, it is sufficient to determine the *number* of customers to contact $|C_t|$. We use this reduced DP formulation in our numerical experiments. The DP is solved by backward recursion, storing the optimal number of customers to contact for all numbers of customers yet to be contacted and all future contacting points.

3.4.2. Attended Home Delivery

In comparing the contacting setting in the AHD and C&C cases, we observe that the routing component significantly complicates the problem. Even in the simpler settings of deterministic customer arrivals or with a high number of contacting periods in the AHD setting, it is no longer apparent which customers to contact due to routing interdependence. To nevertheless solve this complex problem for realistic instances, we propose a three-step procedure. In the first step, the total delivery area is decomposed into subareas. In the second step, it is determined which of these subareas should be served, and in the third step, the decisions regarding which customers to contact are made. The following subsections discuss the individual steps in more detail.

3.4.2.1. Subdivision of Delivery Area into Subareas

Since the problem cannot be solved directly for the entire delivery area due to its high complexity, the delivery area is first decomposed into smaller subareas under the assumption that a single van serves a corresponding subarea by completing one cycle within the area per delivery time slot. This approach to roughly estimating delivery costs was proposed by Daganzo (1987) under a number of assumptions, such as uniform distribution of customers over time slots and subareas. In our considered crisis scenario, these normally fairly strong assumptions are in fact realistic due to the extremely high demand.

We follow a similar clustering procedure as in Yang and Strauss (2017) who likewise built a geographic decomposition approach on Daganzo (1987). To accomplish the subdivision, the total delivery region is first partitioned into bands (north to south) of equal size. Since the heights of the rectangles are determined by the conveniently chosen number of horizontal bands, in a second step, only the horizontal widths (west to east) of the rectangles need to be determined. Two requirements must be considered in the process of determining the horizontal widths. First, there should not be too many customers in the rectangles (due to the maximum physical capacity of the vehicles), and second, the rectangles should not be geographically too large, depending on the distance from the depot (due to the maximum length of the tour). To ensure that the rectangles contain the right number of customers, we introduce a limit on the number of customers in the rectangles. We set the limit so that in expectation as many customers arrive as can be supported by physical capacity. This limit can be calculated as $\frac{MaxCap}{\lambda}$. To ensure that the maximum length of the tour is not exceeded, an upper bound on the geographical size of the rectangle, depending on the distance to the depot, must be considered. We make use of the "cluster-first, route-second" approach that was introduced in Daganzo (1987). Accordingly, the average distance traveled by the vehicle from the depot to the rectangle, within the rectangle, and back to the depot can be calculated as

$$2\rho + l + \frac{wE}{6} + \tau E \leq MaxLength,$$

where ρ describes the distance from the depot to the center of the rectangle, l and w denote the length and width of the rectangle, E is the expected number of arriving customers in the rectangle and τ is the dwell time at the customer location. For further information on this formula, we refer to Daganzo (1987).

To allow a differentiation between customer segments, we develop this rectangular structure for different combinations of customer segments. Here, we proceed hierarchically, starting with the highest-weighted customers. Thus, we start with customers from Segment 1 only and determine corresponding rectangles. Subsequently, we add the customers from Segment 2 and determine new rectangles with the pool of customers from Segments 1 and 2. We follow this procedure until we finally consider the total set of customers. In this way, we obtain as many different sets of rectangles as there are customer segments. Combining all rectangles, we obtain the overall set of overlapping rectangles $\mathcal{R}$.

3.4.2.2. Selection of Subareas

Once the total delivery area is decomposed into smaller subareas using the method described in the previous subsection, in a second step, it must be determined which of these overlapping areas are most favorable to be served. Consequently, a certain subset of subareas needs to be selected that is served by the vehicles. To find the most promising non-overlapping subareas, we need to solve the following selection model. The model includes the binary decision variable $x_r \in \{0,1\}$ for every rectangle $r \in \mathcal{R}$, indicating whether the rectangle r will be served:

$$
\begin{aligned}
\max \sum_{r \in \mathcal{R}} & m_r \cdot x_r \\
\sum_{r \in \mathcal{R}} & x_r \leq \mathcal{V} \\
& x_s + x_t \leq 1 && \forall\,(s,t) \in \mathcal{O} \\
& x_r \in \{0,1\} && \forall\, r \in \mathcal{R}.
\end{aligned}
$$

Since we assume that all customers, regardless of their segment, arrive with the same probability if contacted, we maximize the expected total mass of the rectangles. The mass of a rectangle $r \in \mathcal{R}$ is denoted by $m_r = \sum_{c \in C(r)} \alpha_c$ as the sum of weights α_c of all customers $c \in C(r)$ in rectangle r. The first constraint of the model ensures that the total number of chosen rectangles does not exceed the number of vehicles. The second class of constraints guarantees that the chosen rectangles do not overlap, i.e., no customer is contacted twice. Here, the set $\mathcal{O}$ consists of all pairs (s,t) of rectangles $s,t \in \mathcal{R}$ with $C(s) \cap C(t) \neq \emptyset$.

3.4.2.3. Area-Based Contacting Approaches for Attended Home Delivery

After determining which of the rectangles to serve, the question arises of which customers in the subareas must be contacted in which contacting period. In the following paragraphs, we discuss several contacting approaches. First, we consider the simple approach of contacting all customers in the selected rectangle. Then, we investigate the case of deterministic customer arrivals. Last, we investigate a more sophisticated contacting approach that is based on the contacting concept used in the C&C setting.

Contacting All Customers of the Selected Subareas
A simple option for area-based customer contacting would be to contact all customers from the selected rectangles at once. Since all customers are contacted and then served according to the FCFS principle, the entire contacting problem reduces to a one-period problem.

Contacting with Deterministic Customer Arrivals in the Selected Subareas
If we assume that contacted customers arrive with certainty, the problem, like the deterministic case in the C&C setting, reduces to a one-period problem. Because of deterministic customer arrivals, no overbooking occurs, and only the question of which customers to serve from the selected rectangles remains. This problem is similar to the problem considered in Cleophas and Ehmke (2014) as there is no difference whether customers are proactively contacted and arrive with certainty or whether (predicted) arriving customers are accepted. However, since we want to use this special case as a benchmark in our numerical study, we utilize the TOP to solve the problem instead of applying an acceptance algorithm as in Cleophas and Ehmke (2014) or Ehmke and Campbell (2014). In particular, we use the TOP formulation from Tang and Miller-Hooks (2005) and adapt it to our specific problem setting. The formulation starts with a complete graph $G = (V, E)$, where $V = \{0, \dots, n\}$ describes the vertex set, consisting of the depot (vertex 0) and the potential customers to be served from the selected rectangles, $E = \{(i, j) \,|\, i, j \in V\}$ describes the edge set and α_i denotes the weight collected by visiting customer i for $i \in V \setminus \{0\}$. The idea is to find a set of m vehicle tours that start and end at the depot and collect the maximum total weight by visiting the selected vertices within a maximum tour length $MaxLength$ and respecting the maximum physical capacity $MaxCap$ of the vehicles. The distance between i and j for $i, j \in V$ is denoted by d_{ij}, and we use the following decision variables:

- $y_{ik} \in \{0, 1\}$: vertex $i \in V \setminus \{0\}$ is visited by vehicle $k \in \{1, \dots, m\}$,
- $x_{0jk} \in \{0, 1, 2\}$: number of times edge $(0, j)$ for $j \in V \setminus \{0\}$ is traversed by vehicle $k \in \{1, \dots, m\}$,
- $x_{ijk} \in \{0, 1\}$: number of times edge (i, j) for $i < j$, $i, j \in V \setminus \{0\}$ is traversed by vehicle $k \in \{1, \dots, m\}$.

With this notation, we formulate the TOP as

$$\max \sum_{i=1}^{n} \sum_{k=1}^{m} \alpha_i y_{ik} \tag{3.1}$$

$$\sum_{j=1}^{n} \sum_{k=1}^{m} x_{ojk} = 2m \tag{3.2}$$

$$\sum_{i<j} x_{ijk} + \sum_{i>j} x_{jik} = 2y_{jk} \qquad \forall\, j \in \{1, ..., n\},\, k \in \{1, \ldots, m\} \tag{3.3}$$

$$\sum_{i=0}^{n-1} \sum_{j>i} d_{ij} x_{ijk} + \sum_{i=1}^{n} \tau y_{ik} \leq MaxLength \qquad \forall\, k \in \{1, \ldots, m\} \tag{3.4}$$

$$\sum_{k=1}^{m} y_{ik} \leq 1 \qquad \forall\, i \in \{1, \ldots, n\} \tag{3.5}$$

$$\sum_{i<j,\, i,j \in U} x_{ijk} \leq |U| - 1 \qquad \forall\, U \subset V \setminus \{0\},\, n-1 \geq |U| \geq 2,\, k \in \{1, \ldots, m\} \tag{3.6}$$

$$\sum_{i=1}^{n} y_{ik} \leq MaxCap \qquad \forall\, k \in \{1, \ldots, m\} \tag{3.7}$$

$$x_{0jk} \in \{0, 1, 2\} \qquad \forall\, j \in \{1, \ldots, n\},\, k \in \{1, \ldots, m\} \tag{3.8}$$

$$x_{ijk} \in \{0, 1\} \qquad \forall\, 1 \leq i < j,\, j \in \{1, \ldots, n\},\, k \in \{1, \ldots, m\} \tag{3.9}$$

$$y_{ik} \in \{0, 1\} \qquad \forall i \in \{1, \ldots, n\},\, k \in \{1, \ldots, m\}. \tag{3.10}$$

The objective is to maximize the overall sum of the collected weights of the customers who are served. The first constraint initializes the tours of the m vehicles. The second class of constraints guarantees the connectivity of the vehicle tours and links the variables x and y. The next class of constraints restricts the maximum lengths of the vehicle tours. Constraints (3.5) ensure that a customer can be served at most once, while constraints (3.6) prohibit sub-tours. The last class of constraints ensures that the maximum physical capacity of the vehicles is not exceeded. Using this model, the special case of proactive customer contacting with deterministic customer arrivals in the selected subareas can be solved.

Targeted Customer Contacting

If we assume stochastic customer arrivals, a more sophisticated procedure of proactively contacting customers can be derived from the insights of the contacting procedure in the C&C setting. Since the rectangle formation already addresses and constrains the geographical component of this problem, the underlying concept is to apply the DP formulation of proactively contacting customers from the C&C setting (Section 3.4.1) in the selected subareas. Thus, the overall AHD problem is interpreted as the combination of multiple C&C problems. However, it should be noted that the customer location information must also be included in the contacting process. In the C&C setting, we observed that in certain instances only some customers of a given priority group are contacted. In C&C, it is irrelevant which of the customers are selected from the priority group because the customers are of equal value and lead to the same resource consumption. Since the locations of the served customers influence the routing decision, in this approach, we contact the customers in a weight-decreasing, distance-to-depot-increasing sequence. This choice ensures that the routing procedure is as flexible as possible.

3.5. Numerical Study

In this section, we explain the design of our numerical study, report our results, and discuss ensuing managerial insights.

3.5.1. Click&Collect

To evaluate the benefits of demand management through proactive customer contacting in the C&C setting in the event of a crisis, we test our proposed approach on a realistic data set. With these experiments, we seek to answer the following questions: What is the benefit of proactive customer contacting for an e-grocer offering the Click&Collect fulfillment method in times of crisis? What parameters drive this benefit? What are the costs incurred by the e-retailer to prioritize vulnerable customers in times of crisis?

As a benchmark for our proposed approach, we implement the FCFS approach, which is currently used by our industry partner. In addition, we also calculate the asymptotic results for our contacting approach with a large number of contacting periods. Our numerical study thus includes the following tested policies:

- FCFS: no active demand management, customers arrive simply according to FCFS until capacity is depleted.
- Contacting: proactive customer contacting approach with different response probabilities λ of customers and different numbers T of contacting periods.

3.5.1.1. Data and Parameters

Before presenting the results of our study, we explain the relevant parameter settings summarized in Table 3.1. The data used in this numerical study can also be accessed online (Schwamberger et al. 2022a). The considered scenario reflects the setting in the region of London, UK. All values are either provided by our industry partner, a well-known supermarket chain that operates C&C and AHD services all over the UK, or, if no information was provided, estimated based on publicly available data.

The capacity of customers that can be served by the C&C service per day per pick-up store is limited by the store operations and was given by our industry partner as $K = 100$. To achieve customer prioritization, we consider $L = 3$ customer segments: Segment 1, which contains the most preferred customers, includes all customers from the EDV list;

Table 3.1.: Parameters C&C

Parameter	Meaning	Values
K	Capacity	100
L	Number of customer segments	3
$\mathcal{C}$	Number of customers per segment	[48; 54; 375]
α	Weights of customer segments	[10; 7; 5]
λ	Response probability	0.9 / 0.75 / 0.6
T	Contacting periods	1 / 2 / 3 / 4 / 5 / 6
β	Penalty factor	3
r	Retailer's revenue per segment	[43.5; 61; 61]

Segment 2 includes customers owning a premium pass for online grocery orders; Segment 3 consists of the remaining customers who do not receive special prioritization. To assess the relative sizes of the segments, we estimate the total number of customers using the C&C service. Using the total number of inhabitants of London, the percentage of people who buy groceries online and the market share of our industry partner, we obtain approximately 400,000 customers in the London area. With this number of customers, as well as the information pertaining to their age distribution, the respective shares of age groups that buy groceries online and the share of customers who own a premium pass, we are able to determine the relative sizes of the segments. In total, we obtain 40,000 customers in Segment 1 ($\approx 10\%$), 45,000 customers in Segment 2 ($\approx 11.25\%$) and 315,000 customers in Segment 3 ($\approx 78.75\%$). Since approximately one quarter of these online customers use the C&C service, our industry partner operates 30 stores in the London area, and assuming that our customers purchase groceries once per week, we obtain a total number of 48 customers for Segment 1, 54 customers for Segment 2 and 375 customers for Segment 3 per store per day. The weights reflecting the prioritization of these segments are set as $\alpha_1 = 10$, $\alpha_2 = 7$, and $\alpha_3 = 5$. We vary the response probability $\lambda = 0.9/0.75/0.6$ of customers responding to the contacting, and the number of contacting periods. Assuming that customers are given 12 hours in which to respond after being contacted and that the contacting phase begins at most three days before the cut-off time, the following numbers of contacting periods $T = 1/2/3/4/5/6$ seem appropriate. If a customer is contacted and cannot be served, penalty costs of $\beta \cdot \alpha$ are incurred. We set $\beta = 3$, which corresponds to a reasonable Newsvendor fractile of 75%. To quantify the impact of this prioritization process on the retailer's revenue, we estimate the expected revenue per segment. Since the average weekly spending on food of a person aged 75 years or older is £43.5, significantly lower than the average weekly expenses of £61 (Statista 2018), we assume realized revenues of £43.5 for a customer from Segment 1 and £61 for customers from Segment 2 or Segment 3.

3.5.1.2. Results of the Numerical Study

We first address the relative benefits of contacting versus FCFS. Table 3.2 provides a summary of the numerical results and is discussed in this section. The third column of this table describes the overall expected score of each policy in the different settings. The fourth column presents the score compared to the FCFS method in %. The fifth column describes the score on a range of 0-100% relative to the gap between the FCFS method and the contacting setting with an infinite number of contacting periods. The following columns show the numbers of customers to contact in the first contacting period and the expected numbers of served customers over the entire contacting horizon for each customer segment. Then, following the expected capacity utilization in %, the expected share of rejected customers out of the arriving contacted customers is given in %, and the expected average revenue per served customer is given in £. All computations were conducted using an HP ProBook 440 G5 with a 2.5 GHz Intel Core i5 processor (two cores) with a run time of less than two hours per contacting period for these instances. Note that the contacting problem can be solved offline before customers are contacted.

Results FCFS

Since the considered set of customers is several times larger than the available capacity, the customers' response probability has no influence on the results, as the demand exceeds capacity for all response probabilities. Therefore, Table 3.2 includes only one line for the FCFS approach. As the response probability is identical across all customer segments and the customers are served until capacity is depleted (capacity utilization of 100%), all customer segments receive the same service level.

Results Contacting

The results for the proactive contacting approach with stochastic customer arrivals are listed by decreasing response probabilities and increasing number of contacting periods. Note that for every T at the beginning of each contacting process, it is assumed that the entire capacity and finite number of customers are available.

Table 3.2.: Numerical Results: Contacting for C&C

	T	Score	Comp. with FCFS (%)	FCFS $\leftrightarrow \infty$ (%)	Cust. to cont. in T Seg. 1	Cust. to cont. in T Seg. 2	Cust. to cont. in T Seg. 3	Exp. served cust. Seg. 1	Exp. served cust. Seg. 2	Exp. served cust. Seg. 3	Cap. (%)	Rej. (%)	Rev. (£)
FCFS		572.94	100.00	0.00				10.06	11.32	78.62	100.00		59.24
Contacting ($\lambda = 0.9$)	1	790.22	137.92	90.44	48	54	6	43.08	48.46	5.38	96.93	0.29	53.22
	2	809.80	141.34	98.58	48	54	2	43.20	48.60	7.58	99.37	0.00	53.31
	3	812.78	141.86	99.83	48	54	1	43.20	48.60	8.13	99.93	0.00	53.44
	4	813.15	141.93	99.98	48	53	0	43.20	48.60	8.19	99.99	0.00	53.44
	5	813.19	141.93	100.00	48	52	0	43.20	48.60	8.20	100.00	0.00	53.44
	6	813.20	141.93	100.00	48	52	0	43.20	48.60	8.20	100.00	0.00	53.44
	∞	813.20	141.93	100.00	48	52	0	43.20	48.60	8.20	100.00	0.00	53.44
Contacting ($\lambda = 0.75$)	1	724.78	126.50	80.74	48	54	25	35.85	40.33	18.67	94.85	0.42	54.38
	2	752.02	131.26	95.22	48	54	19	35.99	40.48	22.36	98.83	0.18	54.63
	3	757.84	132.27	98.32	48	54	16	36.00	40.50	23.03	99.52	0.04	54.67
	4	760.00	132.65	99.47	48	54	14	36.00	40.50	23.32	99.82	0.01	54.69
	5	760.71	132.77	99.85	48	54	13	36.00	40.50	23.45	99.95	0.00	54.70
	6	760.92	132.81	99.96	48	54	11	36.00	40.50	23.49	99.99	0.00	54.70
	∞	761.00	132.82	100.00	48	52	0	36.00	40.50	23.50	100.00	0.00	54.70
Contacting ($\lambda = 0.6$)	1	663.75	115.85	66.84	48	54	55	28.63	32.21	32.81	93.64	0.58	55.65
	2	695.64	121.42	90.31	48	54	46	28.78	32.34	37.17	98.33	0.32	55.88
	3	702.77	122.66	95.56	48	54	42	28.79	32.39	38.07	99.26	0.15	55.92
	4	705.72	123.18	97.73	48	54	39	28.80	32.40	38.37	99.57	0.05	55.94
	5	707.27	123.45	98.87	48	54	37	28.80	32.40	38.57	99.76	0.02	55.95
	6	708.08	123.59	99.47	48	54	36	28.80	32.40	38.68	99.88	0.00	55.96
	∞	708.80	123.71	100.00	48	52	0	28.80	32.40	38.80	100.00	0.00	55.96

First, consider the results for the highest response probability ($\lambda = 0.9$). Even if only one contacting point is available, the achieved score is already more than 37% higher than that for the FCFS approach. This is the case even though capacity utilization is slightly lower (96.93%) and a small fraction of customers must be rejected (0.29%). The rationale is that the numbers of customers served from Segments 1 and 2 have increased significantly. This is a direct result of the number of contacted customers. When the number of contacting points increases, the achieved scores increase further. However, it is notable that the score increases only marginally after approximately four contacting points. At this point, the obtained score is close to the upper bound (i.e., $T = \infty$). At this number of contacting points, no customers are rejected, and the capacity is almost completely utilized. This is due to the increased number of customers being served from Segment 3. In fact, fewer customers from Segment 3 are contacted in the first contacting point. However, since the e-retailer has more opportunities to observe how many of the contacted customers arrive, capacity utilization can be increased.

With decreasing response probabilities, an overall decreasing score can be observed, resulting from deteriorating market potential. This can be seen in the increasing proportion of customers served from Segment 3, although similar numbers of customers are contacted from Segment 1 and 2. In general, however, similar effects for the score, capacity utilization and share of rejected customers to the response probability $\lambda = 0.9$ are observable. It is noteworthy that even with these response probabilities, approximately 4 contacting points are sufficient, although the overall convergence to the asymptotic score is not as rapid. Overall, it becomes apparent that the number of overbookings, i.e., the aggressiveness of contacting, increases.

Table 3.2 also documents the cost of prioritizing vulnerable customers. Serving more customers from Segments 1 and 2 decreases the average revenue per customer served, albeit only slightly.

Managerial Insights: Click&Collect

Demand can be strongly controlled by proactively contacting customers, which answers the first question. Comparing the numbers of the various response probabilities to the initial situation, a significant increase in the number of served customers from Segments 1 and 2 can be observed at all response probabilities. Despite a minor drop in capacity utilization, this has a positive effect on the total score, which improves significantly in all cases compared to the status quo. In particular, we identify great improvement potential for high response probabilities. This shows that proactively contacting customers is an effective demand management tool which allows the retailer to concentrate on satisfying demand from the most preferred customer groups.

To answer the question of which parameters drive the benefits of proactive customer contacting, it becomes apparent that the aggressiveness of the contacting is closely linked to the response probability and the number of contacting periods available. As the response probability decreases, the uncertainty of how many customers actually arrive once contacted increases. If, in addition, only a few contacting periods are available, the aggressiveness of the algorithm increases. This leads to increasing penalties and, with decreased market potential, to a lower overall score.

Furthermore, it is remarkable how quickly the expected score and the retailer's revenue approach the values of the asymptotic contacting behavior, which becomes evident by their relative comparison to the upper bound. Although the actual contacting behavior observed in period $T = 4$ is still very different from the asymptotic results, there are hardly any significant differences in the resulting key indicators, such as capacity utilization, revenues and total score. The marginal benefit of adding an additional contacting point diminishes quickly and it is sufficient to establish only a few contacting points.

Another managerial insight is that customer prioritization in the C&C setting hardly affects the capacity utilization. In all tested instances, a capacity utilization of at least 99% is achieved with only three contacting points.

Retailers face a trade-off between prioritizing customers according to their vulnerability and short-term revenues. This trade-off can be observed in the results and is much milder than originally expected. In our setting, at least a tripling of customers served from Segments 1 and 2 can be achieved by accepting a decrease in average revenue per customer served by at most 10%. Note that this trade-off is problem inherent and is not due to the customer contacting. If the retailer values revenues more highly, the segment weights can be adjusted accordingly.

Summarizing, the gained managerial insights are:

- Proactive customer contacting is a valuable demand management tool.
- A high customer response rate benefits the contacting policy; thus, it is particularly suitable in crisis situations.
- The marginal benefit of adding an additional contacting possibility decreases quickly.

- The improvement in the expected customer mix is at the expense of only a minor drop in capacity utilization.
- The trade-off between prioritizing vulnerable customer groups and revenue is relatively mild.

3.5.2. Attended Home Delivery

In addition to the same research questions that we have formulated in Section 3.5.1 in the C&C context, we aim to explore structural differences in the results of proactively contacting customers between the AHD and C&C fulfillment methods with this numerical study.

To obtain a relevant benchmark against which to compare our proposed methods, we use, similar to the C&C setting, the FCFS principle for the entire delivery area. Thereby, the policies tested in this numerical study are as follows:

- FCFS: no active demand management; customers arrive simply according to FCFS from the entire delivery area.
- ContactAll: proactively contacting all customers of the selected subareas.
- ContactTargeted: proactive customer contacting within the selected subareas based on the C&C approach.

In addition, we report the results obtained by first sampling which customers of the chosen subareas arrive, depending on the response probability, and then solving the deterministic TOP on this subset of customers. For given customer arrivals, this provides an ex post upper bound for the contacting procedure which naturally avoids any customer rejections. We label the results as Ex Post TOP.

Before we describe the specific parameters and results of the numerical study, we should highlight a few differences with the numerical study of the C&C setting. Since we assume a fixed capacity and there are no connections between arriving customers in the C&C setting, we are able to calculate expected values directly. In the AHD setting, in contrast, different routes are created depending on the actual arriving customers, thus requiring simulations. Because the study can be parallelized very well after determining the rectangles and selecting the most promising rectangles, the numerical study can be implemented efficiently.

3.5.2.1. Data and Parameters

All parameters that are described in the following are listed in Table 3.3. With the same reasoning as in the previous numerical study for the C&C setting, we assume that there are three customer segments ($L = 3$) with associated weights $\alpha = [10, 7, 5]$ and revenues $r = [43.5, 61, 61]$. In this study, we again investigate three different response probabilities of the customers, reflected by $\lambda = 0.9$, $\lambda = 0.75$ and $\lambda = 0.6$. Once again, the penalty term is set to $\beta = 3$. In addition, we investigate the same numbers of contacting possibilities $T = 1, \ldots, 6$ as in the C&C setting.

Table 3.3.: Parameters AHD

Parameter	Meaning	Values
L	Number of customer groups	3
α	Weights of customer groups	[10; 7; 5]
r	Retailer's revenue	[43.5; 61; 61]
λ	Response Probability	0.9 / 0.75 / 0.6
β	Penalty term	3
T	Number of periods	1 / 2 / 3 / 4 / 5 / 6
$\mathcal{V}$	Number of Vehicles	160
C	Finite set of customers	[1,071; 1,206; 8,437]
$NumberofSimulations$	Number of Simulations	100
$MaxCap$	Max. physical capacity of vehicle	14
$MaxLength$	Max. Length of tour	200
τ	Dwell Time at the customer location	10
$DeliveryArea$	Delivery area	40 x 40
$NumberHorizontalBands$	Number of horizontal bands	15

Compared to the numerical study of the C&C approach, the number of customers $\mathcal{C}$ and the capacity alter since the distribution of customers using C&C and AHD is different. As we have already described in the numerical study of C&C, we have approximately 400,000 customers buying groceries online in the London area. Since approximately three-quarters of these customers use the AHD service, there are 300,000 customers, of which 30,000 belong to Segment 1, 33,750 to Segment 2, and 236,250 to Segment 3. We assume an equal distribution of these customers over the total delivery area. Our industry partner operates one dedicated fulfillment center with 160 vehicles that exclusively fulfills AHD orders. This center with its associated vehicles has a weekly capacity to fulfill 25,000 online orders. Since a van usually operates two shifts per day (7:00 - 15:00 and 15:00 - 23:00) 7 days per week, on average, 11 customers are served per tour. In addition to the dedicated fulfillment center, AHD orders are fulfilled by 30 additional stores. Since a store operates 5 vans on average, with each scheduled to work two shifts 7 days per week, and on average 11 customers are served per tour, we estimate that a store can fulfill approximately 770 orders per week. Combining the capacity from the center and the stores, we conclude that there is the capacity to serve approximately 50,000 orders per week in the London area. Since we want to focus on the requests handled by the dedicated center in one shift, we obtain by assuming a uniform distribution of customer requests between the stores and the dedicated center a total number of 150,000 customer requests per week. Since the vehicles operate two shifts on 7 days, this corresponds to a total number of 10,714 customer requests per shift. With the relative sizes of the customer segments, this results in approximately 1,071 customers from Segment, 1,206 customers from Segment and 8,437 customers from Segment 3.

Since explicit routes need to be planned in the AHD setting, we performed simulations for this numerical study. We fix the number of simulations at 100. Additionally, the vans used to perform the routes have a certain physical capacity. Since a van can carry 70 crates (ambient crates, cooled crates and frozen crates) and an average customer order consists of 5 crates, a maximum of 14 customers can be served per tour. Since a delivery shift is 8 hours and the average delivery speed of a vehicle is assumed to be 25 km per hour, the overall tour length is limited to 200 km. In addition to the maximum tour length, a dwell time of 10 minutes is incurred per customer. As the London area contains approximately 1,600 km^2, we assume an underlying delivery area of 40 km x 40 km. Since the depot of our industry partner is located in the vicinity of the city center, we assume that the depot is located in the center of the delivery area. Because we follow the clustering approach of Yang and Strauss (2017) to divide the London area into subareas in our proposed method, we use the identical number of horizontal bands (15) to subdivide the entire delivery area into subareas.

3.5.2.2. Results of the Numerical Study

The results we summarize in the following paragraphs are gathered in Tables 3.4 and 3.5. Table 3.4 provides the average values of the different tested contacting policies with four contacting points for different response probabilities rounded to two decimals. In contrast, Table 3.5 displays the data of the targeted contacting approach for different customer response probabilities and numbers of contacting periods. Similar to the

C&C setting, the third columns of the tables denote the average total score per vehicle. The fourth columns show the average score compared to the FCFS method. The fifth columns denote the score on a range of 0-100% compared to the FCFS method (0%) and Ex Post TOP (100%). Subsequently, in the next columns of Table 3.4 follow the respective average numbers of contacted customers in the first contacting period and the average number of total served customers over the entire contacting horizon per customer segment. Next, both Tables 3.4 and 3.5 provide the average physical capacity and routing utilization in %, the average share of rejected customers from the arriving customers in % and the average revenues per served customer in £. All results presented in both tables are statistically stable and significant. Since a sufficient number of simulations was performed, all scores obtained by the contacting methods are within a 95% confidence interval of at most ± 0.25. The run time to determine the targeted contacting policy was less than 30 seconds for each subarea. As in the C&C setting, the contacting problem can be solved in advance. The only online problem that needs to be solved is whether an arriving customer can be added to the existing route. Using this heuristic, this problem can be solved in a fraction of a second for each subarea.

Results FCFS

Before diving deeper into the results of our proposed contacting procedure, we first consider the results of the FCFS method on the entire delivery area (Table 3.4). Since we assume that demand is several times larger than the available capacity, the specific response probabilities of the customers are irrelevant, resulting in only one line for the FCFS approach in Table 3.4. As in the C&C setting, all customer segments obtain the same service level, and only the relative sizes of the customer segments determine the achieved score. When serving the customers in a FCFS sequence, the routes are not planned very efficiently since the customers are simply inserted based on the nearest insertion heuristic (Section 3.3.2). Both the average routing and the average physical capacity utilization are high. However, compared to the C&C setting, the physical capacity is not completely exploited at 95%.

Table 3.4.: Numerical Results: Contacting for AHD (4 Contacting Points)

λ	Method	Score	Comp. with FCFS (%)	FCFS ↔ Ex Post TOP (%)	Cust. to cont. in T			Av. served cust.			Cap. (%)	Rout. (%)	Rej. (%)	Rev. (£)
					Seg. 1	Seg. 2	Seg. 3	Seg. 1	Seg. 2	Seg. 3				
	FCFS	76.43	100.00	0.00				215	239	1679	95	97		59.23
	ContactAll	92.84	121.47	48.04	1063	1192	300	911	1020	256	98	48	4.88	53.71
0.9	ContactTargeted	110.54	144.63	99.85	1063	925	252	957	977	252	98	48	0.00	53.34
	Ex Post TOP	110.59	144.69	100.00				956	980	252	98	48	0.00	53.35
	ContactAll	80.89	105.84	17.46	1066	1188	783	751	835	552	96	47	6.17	54.85
0.75	ContactTargeted	101.68	133.04	98.83	1065	655	520	799	795	540	95	47	0.00	54.45
	Ex Post TOP	101.98	133.43	100.00				804	795	539	96	47	0.00	54.42
	ContactAll	67.89	88.83	-49.02	1063	1184	1587	587	655	876	95	47	8.20	56.15
0.6	ContactTargeted	93.42	122.23	97.53	1062	576	911	639	618	849	94	47	0.03	55.69
	Ex Post TOP	93.85	122.79	100.00				635	626	856	95	47	0.00	55.75

Table 3.5.: Numerical Results: Contacting for AHD

λ	**T**	**Score**	**Comp. with FCFS (%)**	**FCFS $\leftrightarrow$ Ex Post TOP (%)**	**Cap. (%)**	**Rout. (%)**	**Rej. (%)**	**Rev. (£)**
0.9	1	103.27	135.12	78.57	90	46	0.04	52.71
	2	110.26	144.26	99.03	97	48	0.00	53.34
	3	110.47	144.54	99.65	98	48	0.00	52.65
	4	110.54	144.63	99.85	98	48	0.00	53.34
	5	110.58	144.68	99.97	98	48	0.00	53.35
	6	110.58	144.68	99.97	98	48	0.00	53.35
0.75	1	91.27	119.42	58.08	89	45	1.50	54.06
	2	99.93	130.75	91.98	94	47	0.09	54.35
	3	101.61	132.95	98.55	95	47	0.00	54.43
	4	101.68	133.04	98.83	95	47	0.00	54.45
	5	101.73	133.10	99.02	95	47	0.00	54.46
	6	101.73	133.10	99.02	95	47	0.00	54.45
0.6	1	80.81	105.73	25.14	84	44	1.66	55.20
	2	89.84	117.55	76.98	92	46	0.71	55.59
	3	92.45	120.96	91.96	93	47	0.13	55.64
	4	93.42	122.23	97.53	94	47	0.03	55.69
	5	93.67	122.56	98.97	94	47	0.00	55.73
	6	93.8	122.73	99.71	94	47	0.00	55.73

Results Contacting All Customers in the Selected Subareas
If all customers in the selected subareas are contacted (Table 3.4), we observe that reasonable scores can be achieved for the high and moderate response probabilities. If the response probability λ decreases, the variance of arriving customers increases. With an underlying rectangular structure that depends on the response probabilities and the concept of contacting all customers in each rectangle, this often leads to customers who are contacted but cannot be served due to insufficient capacity (see share of rejected customers). At high and moderate response probabilities, the score of this method, including incurred penalties for contacted customers who cannot be served, exceeds the score of the FCFS approach by at least 5%. For a low response probability, contacting all customers in the selected rectangles leads to a slightly inferior overall score, compared to the FCFS approach, because the share of rejected contacted customers - and thus the arising penalties - are quite high. A thoroughly high physical capacity utilization indicates that the first step of the overall procedure, the division of the entire delivery area into smaller subareas, is an appropriate way to geographically decompose the problem. In all settings, a capacity utilization of at least 95% is achieved (similar to the FCFS method), meaning that every vehicle, on average, has less than one spare unit of capacity. As the numbers of served customers indicate, the method of contacting all customers in the selected subareas can already be used to steer the demand effectively. For example, the number of customers served from Segments 1 and 2 is more than quadrupled for $\lambda = 0.9$ compared to the FCFS approach. However, a direct comparison with Ex Post TOP shows that there is still room for improvement.

Results Targeted Customer Contacting
Our more sophisticated contacting approach based on the C&C contacting procedure uses the possibility of additional contacting periods and waits to observe how many of the contacted customers actually arrive. This limits the number of overbookings and yields better overall scores. For the sake of simplicity, only the values for four contacting periods are displayed for the targeted contacting method in Table 3.4. As seen in the table, a high score is obtained for a high response probability $\lambda = 0.9$ compared to the FCFS principle. The lower the response probability is, the lower the score, both compared to the results of the FCFS method and on the range between FCFS and Ex Post TOP. This is consistent with the results from the C&C setting and confirms, together with the indicated number of served customers per customer segment, that

in the AHD setting, the lever of proactively contacting customers can also be used to effectively manage demand, especially if the customers are desperate to place an order. In contrast to the C&C setting, however, we observe that the overall capacity utilization is slightly lower in the AHD setting. This is due to situations in which all customers in the subareas are contacted (see the previous method), but too few customers appear to exploit the complete capacity. This could easily be remedied with a further increase in the size of the subareas. However, given the consistently high capacity utilization of at least 94%, a further inflation of the rectangles provides only marginal improvement. Furthermore, the significantly lower routing utilization of the contacting approaches compared to the FCFS method indicates that the routes can be designed much more efficiently for all response probabilities, given subareas of this size. The average revenues per served customer show a similar pattern as in the C&C case. The lower the response probability, the higher the average revenue per served customer, since on average, more customers with higher revenues (i.e., from Segments 2 and 3) arrive. Since the targeted contacting approaches consist of four contacting opportunities in this table, almost no customers need to be rejected for all response probabilities.

Table 3.5 complements the results by displaying the influence of the number of contacting periods for different response probabilities. We again observe similar patterns as in the C&C setting. With decreasing response probabilities, an overall decreasing score can be observed, that results from the deteriorating market potential. An increasing number of contacting periods increases the performance of the contacting approach. This can be detected by the increasing score and the increasing capacity utilization with an increasing number of contacting periods. As in the C&C setting, the improvement levels off quickly. This means that also in the AHD case, a moderate number of contacting periods is sufficient to exploit proactive customer contacting.

Managerial Insights: Attended Home Delivery

Summarizing, similar to the C&C setting, it holds for the AHD setting that the higher the response probability is, the more the lever of proactively contacting customers can be exploited. The same observation holds true for the number of contacting periods. By increasing the number of contacting periods, further flexibility in terms of checking the number of customers who actually arrive once contacted can be added to the contacting process, and the contacting aggressiveness decreases. With a small number of contacting periods, this contacting approach can be used effectively. Compared to the C&C setting, it is notable that in the AHD context, physical capacity utilization is slightly lower due to geographical subdivision into subareas. Overall, however, the capacity utilization is still very high, and every vehicle has less than one spare unit of capacity. Another managerial insight is that the routing component of the AHD setting is already reasonably captured in the subdivision of the area into subareas since the maximum physical capacity, and not the maximum routing length, appears to be the bottleneck. This supports the idea of interpreting the AHD problem as a combination of multiple C&C problems in the subareas. We observe, as in the C&C setting, a mild trade-off between prioritizing vulnerable customers and generated revenue. By proactively contacting customers, the number of customers served from Segments 1 and 2 can be approximately tripled, with a maximum reduction in revenue of 10%.

3.6. Post-Crisis Contacting

Our study is motivated by observations during the recent COVID crisis. In particular, the proposed customer contacting procedure addresses a setting in which demand grossly exceeds capacity. In this section, we explore the potential of this approach in non-crisis times.

We start by analyzing how the benefit of proactive customer contacting changes if demand decreases from the crisis-time highs to more regular levels. To this end, we numerically evaluated both the contacting approach (for a single contacting period, i.e., $T = 1$) and the FCFS benchmark for varying levels of the demand-capacity ratio in the C&C setting. Specifically, we varied λ to scale expected demand to given target multiples of total capacity. Other parameter values are identical to Section 3.5.1.

Table 3.6 shows the results of these experiments. We observe that for very low demand-capacity ratios, contacting coincides with the FCFS approach. Due to the low demand level, it is optimal to contact all customers, and no responding customers are rejected. For increasing demand volumes, proactive contacting initially becomes inferior to FCFS. This is due to the risk of costly overbooking which leads the retailer to contact fewer customers, thereby resulting in a lower capacity utilization. The relative performance of contacting versus FCFS reaches its lowest level at a demand-capacity ratio of 100%. For higher ratios, the relative performance of the contacting approach increases again, due to the stronger prioritization of the most valued customer groups. For a demand-capacity ratio of 175%, FCFS and contacting achieve the same scores, however by different means. For higher ratios, the contacting approach is superior.

The described results have two implications. First, the benefit of proactive customer contacting, as opposed to reactive FCFS customer acceptance, depends strongly on the system load. The presented contacting approach addresses a situation of grossly overwhelmed capacity. In our study, switching to 'crisis mode' becomes beneficial roughly at a demand-capacity ratio of 2:1. Naturally, the specific switching point depends on the market composition.

Second, to be effective in non-crisis times, the contacting approach requires modifications. In such a setting, the benefit of proactively contacting customers does not rest on capacity rationing but rather on leveling short-term demand fluctuations. To exploit this potential, contacting does not replace but rather complement the regular order acceptance process. That is, the retailer can decide towards the end of a regular

Table 3.6.: Numerical Results: Post-Crisis Contacting

Exp. Dem./ Cap. (%)	Method	Score	Comp. with FCFS (%)	Cust. to contact			Exp. served cust.			Cap. (%)	Rej. (%)
				Seg. 1	Seg. 2	Seg. 3	Seg. 1	Seg. 2	Seg. 3		
50	FCFS	286.47					5.03	5.66	39.31	50	
	Contact	286.47	100	48	54	375	5.03	5.66	39.31	50	0
75	FCFS	429.73					7.55	8.49	58.96	75	
	Contact	429.56	99.96	48	54	375	7.55	8.49	58.95	74.99	0.01
100	FCFS	572.94					10.06	11.32	78.62	100	
	Contact	514.32	89.77	48	54	341	9.95	11.20	70.71	91.86	1.07
125	FCFS	572.94					10.06	11.32	78.62	100	
	Contact	533.84	93.18	48	54	253	12.45	14.01	65.64	92.1	1.02
150	FCFS	572.94					10.06	11.32	78.62	100	
	Contact	553.48	96.60	48	54	194	14.96	16.82	60.44	92.22	0.93
175	FCFS	572.94					10.06	11.32	78.62	100	
	Contact	573.29	100.06	48	54	152	17.46	19.64	55.29	92.39	0.86
200	FCFS	572.94					10.06	11.32	78.62	100	
	Contact	593.26	103.55	48	54	121	19.96	22.45	50.32	92.73	0.83
225	FCFS	572.94					10.06	11.32	78.62	100	
	Contact	613.44	107.07	48	54	96	22.48	25.29	44.97	92.74	0.70
250	FCFS	572.94					10.06	11.32	78.62	100	
	Contact	633.82	110.63	48	54	77	24.99	28.11	40.08	93.18	0.68

order intake period whether to try and stimulate additional demand to fill remaining capacity. This appears particularly promising for AHD where customers could be contacted that fit well into routes with already accepted customers. A similar idea was, for example, considered by Keskin et al. (2023) in the context of waste collection. Stimulating demand will likely require some price discounts to motivate contacted customers to place an additional order or to place a planned order earlier. The contacting decision thereby faces a new trade-off. While optimal contacting in crisis times balances weighted accepted orders and overbooking penalties, the decision in non-crisis times also has to account for the offered discounts.

Modeling this way of contacting in non-crisis times requires several modifications to the procedure proposed in this paper. First, a single contacting moment per order intake period appears sufficient. Second, the transition probabilities should reflect the arrival of both regular and contacted customers. Third, expected profit becomes the relevant objective function. It is the latter that complicates the analysis since the relevant profits do not only include collected revenues, discounts, and, in the case of AHD, transportation costs but also opportunity costs for demand shifted away from future order intake periods. We see a detailed analysis of proactive customer contacting in non-crisis times as a promising and relevant avenue for future research. We believe that the analysis presented in this paper provides valuable building blocks for this undertaking.

3.7. Conclusion

To conclude, we investigate a new demand management approach for an e-grocer offering C&C and AHD services by proactively contacting customers. We develop a decision policy for the C&C fulfillment method to address the problem of when to contact customers and extend this approach to the AHD setting. To cope with increased problem complexity, we propose a three-step procedure to solve this problem. First, we subdivide the delivery area into smaller subareas; second, we select the most promising subareas; and third, we determine which customers to contact in which contacting period in the chosen subareas. To gain managerial insights and to show the practical benefits of our approaches, we apply both approaches to realistic data from the London area. After developing and investigating our contacting approaches for the C&C and AHD fulfillment methods in times of crisis, we provide decision support when an e-grocer should switch to proactive customer contacting mode and discuss how this new demand management approach can be used in post-crisis times.

As we are the first to investigate proactive customer contacting in the context of e-groceries in times of crisis, not all questions have been ultimately clarified. For example, further investigation could clarify how more complex customer choice behaviors affect the results of this contacting approach. Since our approach assumes that the customers accept all time slots, the rejection of the time windows is only implicitly modelled in the response probability. A more sophisticated customer choice model could be integrated into the contacting approach. This might be a good starting point for future research.

Chapter 4.

Getting the Most out of Online Deliveries - Proactive Customer Contacting in E-Fulfillment[1]

[1]The research presented in this chapter is based on a paper entitled "Getting the Most out of Online Deliveries: Proactive Customer Contacting in E-Fulfillment" (Schwamberger 2023).

Abstract

During the COVID-19 pandemic, some e-grocers began to proactively reach out to certain customers, giving them priority access to online grocery ordering. In this paper, we explore how this concept of proactively approaching customers can be implemented and leveraged in post-crisis times. We propose to integrate proactive customer contacting at the end of the classical booking process to generate additional, good-fitting demand if it is anticipated that not enough inquiries will emerge via the normal booking process. In this paper, we explore under which circumstances it is beneficial to proactively contact customers, how much additional demand to generate and which customers to target. To this end, we propose an optimization formulation to solve the contacting problem and compare the results in a numerical study with simple contacting strategies and a sampled ex post optimal solution to evaluate the merit of proactively contacting customers. We demonstrate that proactive customer contacting can exploit available untapped resources; show that contacting the marginal best customers is not always optimal, as the optimal contacting decision depends on the customers available for contacting; and reveal that none of the simple contacting strategies is universally superior.

4.1. Introduction

For several years, e-grocers have been experiencing a strong growth in demand. In particular, the coronavirus pandemic and related government actions, such as stay-at-home recommendations or imposed lockdowns, have contributed to the growing popularity of ordering groceries online. The unanticipated surge in demand at the beginning of the COVID-19 crisis resulted in the inability to serve all customers due to insufficient capacity. This led some e-grocers to proactively contact customers and give priority access to certain customers (Schwamberger et al. 2022b). Since then, the peak in demand has receded, the capacities have been adjusted, and the acute scarcity situation has been resolved.

As a result, old familiar challenges have resurfaced. A particularly demanding challenge is the uncertainty of customer arrivals and the associated varying utilization of fulfillment processes over several days (Mkansi et al. 2018, Forbes 2022). Many e-grocers struggle with having underutilized delivery capacity on some days while lacking capacity on other days to serve all customer requests. In this paper, we explore how the newly established concept of proactive customer contacting can be leveraged in post-crisis times to address the disparity in fulfillment utilization across multiple days. We propose exploiting this demand management tool in low-demand situations in which an e-grocer anticipates that not enough inquiries will emerge via the normal booking process to entice (future) customers to place an order, leading to higher system utilization and fewer customers to reject on high-demand days. In this regard, we introduce the concept of proactively contacting customers (e.g., via an application or a short text message) as a contingency tool that can be used to shift customer demand in the event of demand scarcity to balance the demand disparity over several days.

Integrating this proactive customer contacting opportunity into the classical booking process (e.g., toward the end of the process, at a strategically determined point) poses several obstacles that must be surmounted. Since the responses of contacted customers are not instantaneous, a decision must be made at the time of contacting regarding whether and which customers should be contacted based on the already accepted orders. Although there is no limit or direct cost of contacting, it is undesirable to contact an excessive number of customers only to reject them. Another obstacle is that the arrival processes for non-contacted and contacted customers run in parallel. This leads to a cannibalization effect, as contacted customers are no longer present in the normal booking process. These challenges lead to various questions that must be addressed to successfully integrate proactive customer contacting in post-crisis times: How many customers should be contacted? Which customers should be target? How does the proactive customer contacting approach affect the achieved profit? What drives the success of proactive customer contacting? We contribute by exploring these questions in this paper. To solve the proactive customer contacting problem, we formulate an optimization problem. We conduct an extensive numerical study in which we compare the optimal contacting solution to simple contacting strategies and a sampled ex post optimal solution to assess the merits of proactively contacting customers in post-crisis times. In addition, we modify the considered contacting setting and evaluate which of the parameters drive the impact of contacting.

The paper is organized as follows: in Section 4.2, we review and discuss the related literature. In Section 4.3, we describe the general problem setting, explain how proactive customer contacting is integrated into the booking process, and formulate an optimization model to solve the problem. Section 4.4 introduces different benchmarking contacting strategies, presents a sampled ex post optimal solution, and provides a numerical study of the contacting approach. Section 4.5 summarizes our results.

4.2. Related Literature

The work presented in this paper interfaces with the research stream of demand management for e-grocery services and the research stream of proactively contacting customers. In recent years, demand management has gained increasing importance in last-mile distribution systems for e-grocery services, e.g., in Agatz et al. (2011), Yang et al. (2016), Yildiz and Savelsbergh (2020), and Fleckenstein et al. (2023). Agatz et al. (2013) identified time slot allocation and time slot pricing as the two main levers of demand management in the context of e-grocers. These decisions are either made on a strategic, tactical or operational level (Wassmuth et al. 2023). Since we focus on a real-time demand management problem in which customers can be contacted to generate additional demand, studies addressing operational time slot offering (Ehmke and Campbell 2014, Mackert 2019) and operational time slot pricing (Yang and Strauss 2017, Klein et al. 2017, 2018) are closest to our work. Wassmuth et al. (2023) provide an in-depth comparison of the individual works in these fields and differentiate the studies depending on the problem setting, the decision-making process and the computational study. Although we likewise use the lever of offering time slots at discounted prices, unlike all these studies, we include the option of proactively contact customers. This results in two parallel arrival processes (of non-contacted and contacted customers), which structurally distinguishes our problem setting from the one usually considered in this literature stream. Our considered setting extends the classical booking process, which gives rise to new research opportunities that constitute the core issue addressed in this paper.

Proactive customer contacting has a solid foundation in direct marketing (Bose and Chen 2009) and customer relationship management (Reinartz and Venkatesan 2008). Typical questions answered in this literature stream include how to profile customers and how to determine customer lifetime values. In our study, however, we do not focus on the customer relationship and marketing issues associated with this contacting process but rather address the optimization issue of which customers to contact in a routing environment. To the best of our knowledge, proactive customer contacting in this regard has only been investigated in Yildiz and Savelsbergh (2020), Keskin et al. (2023), Oetken et al. (2022), and Schwamberger et al. (2022b). Keskin et al. (2023) consider proactive customer contacting in the context of waste collection and investigate the trade-off between more efficient routes and smaller collection volumes resulting from the increased frequency of visits generated by contacting customers. We, in contrast, integrate the proactive customer contacting approach in the context of e-groceries and explore this tool for creating additional demand.

Yildiz and Savelsbergh (2020) consider customer contacting in the context of e-groceries and assume that the time slot pricing decision is the main lever of demand management. In this work, the impact of offering different discounts in exchange for delivery flexibility by contacting customers is investigated. Regarding proactive customer contacting, there are two structural differences compared to our approach. First, customers are proactively approached *after* they have placed an order and offered discounts to reschedule their order to a different time slot that is more convenient for the e-grocer. Second, customers are assumed to respond immediately. In contrast to Yildiz and Savelsbergh (2020), we do not focus on the pricing decision but rather on which customers to contact in the first place, contact only customers who have not yet placed an order to create additional demand, and do not assume instant customer responses.

Recently, Oetken et al. (2022) investigated a similar problem. Similar to Yildiz and Savelsbergh (2020), they assume that the customer contacting process is executed *after* the cut-off time of the normal customer arrival process and that it could even be conducted during the execution of the tour. In contrast to Yildiz and Savelsbergh (2020), it is assumed that the set of regular customers and the pool of target customers are mutually exclusive. As a result of the late contacting decision, contacted customers could only select products from a predefined list of groceries. This is a fundamental difference compared to our approach. Since we assume that proactive customer contacting is executed during the normal booking horizon, we ensure that contacted customers can access all products from the entire assortment, avoiding any dichotomy of customer groups.

The work closest to our manuscript is Schwamberger et al. (2022b), which introduces the idea of proactively contacting customers in crisis situations. However, in that manuscript, the booking process is exclusively reserved for contacted customers due to the fulfillment scarcity situation, and there are multiple contacting opportunities. In this paper, we utilize the idea of proactive customer contacting and integrate customer contacting into the classical booking process in post-crisis times. Since the two arrival processes of contacted and non-contacted customers run in parallel after the unique contacting opportunity, leading to an intricate cannibalization effect, our considered problem is structurally different. Compared to the approach described in Schwamberger et al. (2022b), we also pursue a different objective and maximize the total expected profit.

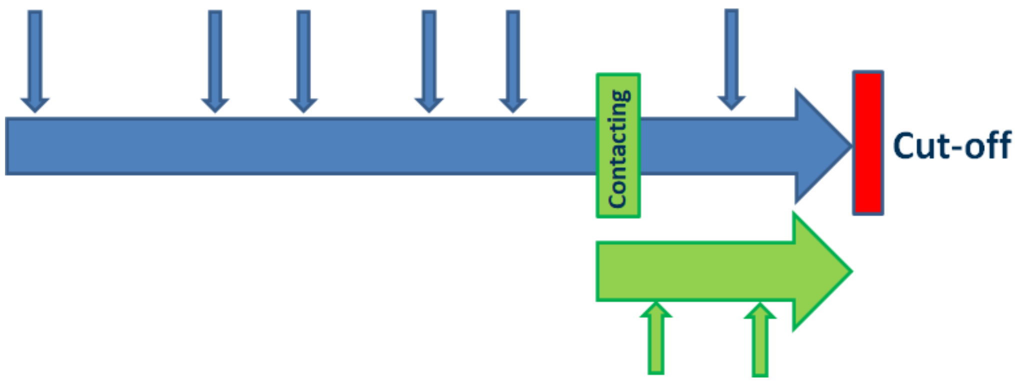
In summary, the analysis presented in this paper draws inspiration from the different studies outlined above. However, no previous research has addressed the problem of an e-grocer seeking to create additional demand by integrating proactive customer contacting into the booking process. Introducing and analyzing this novel planning problem is the way in which our paper contributes to the literature.

4.3. Problem Setting & Model Formulation

We consider an e-grocer with a booking horizon where customer orders arrive dynamically over time. Because customer orders are uncertain over multiple days, the e-grocer struggles with having underutilized delivery capacity on some days while lacking capacity on other days to serve all customer requests. To counteract this disparity over several days, the e-grocer enables proactive customer contacting. For this purpose, the e-grocer establishes a single strategically determined contacting opportunity toward the end of the booking process. To incentivize contacted customers to place an order, a financial discount (e.g., in the form of free delivery) is offered to the contacted customers. Since customers are encouraged to place an order, proactive customer contacting is employed in low-demand situations in which the e-grocer anticipates that not enough inquiries will emerge via the normal booking process.

The problem of deciding whether to proactively contact customers occurs toward the end of an e-grocer's classical booking process, for instance, two hours before the ordering deadline is reached. We assume that this contacting point is strategically determined by the e-grocer. On the one hand, customers should be given a certain time to respond to contacting. On the other hand, the contacting point should not be set too early, as contacting should not be conducted when the majority of orders are still pending (which depends on the e-grocer's previous experience). The e-grocer should consider this trade-off when establishing the contacting point. Figure 4.1 illustrates the booking process with contacting.

Figure 4.1.: Booking Process with Contacting



At the moment of contacting, there already exist accepted customer orders that have to be fulfilled. These customers c_i for $i \in \{1, \dots, N\}$ have known locations $l_i = (x_i, y_i)$ and have to be served in certain time windows $t_i = [a_i, b_i]$. The service time required for serving a customer is denoted by τ. We assume that the delivery area is partitioned into several service regions and that a single vehicle is used to fulfill the customers' orders in each region. Based on the already accepted orders and time slots in a certain region, a preliminary route $\mathcal{R}$ is determined. We assume that the associated routing costs can be calculated using the travel distances and a multiplier γ representing the cost per km. Based on this preliminary route, a decision must be made regarding which set of customers C to contact from the set of available customers to contact $\mathcal{C}$ who have not yet placed an order and can, individually, be feasibly inserted into the already existing route. These customers $c \in \mathcal{C}$, who are available for contacting or may arrive without being contacted, are associated with revenues r_c.

After this unique contacting point, the two arrival processes from contacted and non-contacted customers run in parallel. This leads to a cannibalization effect, as contacted customers might have arrived via the classical booking process without being contacted. We denote the arrival probability of non-contacted customers by $\tilde{\lambda}$ and assume that a contacted customer responds and places an order with a response probability λ. Since the process of proactively approaching customers is performed toward the end of the booking process, we assume that the contacting decision does not affect the timing of placing an order. This ensures that after the contacting opportunity, it is not exclusively the contacted customers who arrive first; rather, the arrival sequence of customers is independent of the contacting status. We assume that customers arriving after the contacting point (regardless of whether they were contacted or not) will be accepted according to the FCFS principle.

Whether an arriving customer, either contacted or not, can be served depends on *two* limiting factors:

1. The maximum number of customers $MaxCap$ that can be served per tour, which is determined by the physical capacity of the vehicle.

2. The time restrictions resulting from the time windows of the already accepted customers and the maximum tour length $MaxLength$ of the vehicle (due to predefined shift lengths and working time restrictions of the employees).

Since checking whether an arriving customer can be served and inserted into the existing route is already a complex routing problem, we use the nearest insertion heuristic to evaluate whether an arriving customer can be added to the preliminary route based on these restrictions. If a non-contacted customer arrives and can be included, a fixed delivery fee δ is collected in addition to the associated revenue r_c (which can be approximated based on a forecast of the customer's order value). If a contacted customer places an order, there is no delivery fee, and only the revenue is obtained. Since proactive customer contacting can shift future demand forward to an earlier date, opportunity costs o_c (for details, see below) are incurred to account for the impact on future routes and customers. In the event of insufficient capacity to serve a contacted customer $c \in C$, this customer has to be rejected, and a penalty β is be incurred.

Considering this problem setting, the objective of the contacting decision is to maximize the expected profit based on the preliminary route at the moment of contacting for the remaining booking horizon. This profit can be calculated as the expected revenue (including the delivery fees of non-contacted customers) minus the additional routing costs of served customers, potential penalties incurred by rejected contacted customers, and opportunity costs of served contacted customers. We have

$$\begin{aligned}
\text{Profit} &= \max_{C \subset \mathcal{C}} \sum_{B \subset \mathcal{C}} \mathbb{P}(B|C) \sum_{\sigma_B \in Perm(B)} \mathbb{P}(\sigma_B) \cdot \text{Profit}(\sigma_B) && (4.1)\\
&= \max_{C \subset \mathcal{C}} \sum_{B \subset \mathcal{C}} \lambda^{|B \cap C|} \cdot (1-\lambda)^{|C \setminus B|} \cdot \tilde{\lambda}^{|B \setminus C|} \cdot (1-\tilde{\lambda})^{|\mathcal{C} \setminus (B \cup C)|} && (4.2)\\
&\cdot \sum_{\sigma_B \in Perm(B)} \frac{1}{|B|!} \cdot \left(rev_C(\sigma_B) - cost(\sigma_B) - pen_C(\sigma_B) - opp_C(\sigma_B) \right) && (4.3)
\end{aligned}$$

where $\mathbb{P}(B|C)$ denotes the probability that the set $B \subset \mathcal{C}$ of customers arrives given that the set $C \subset \mathcal{C}$ of customers is contacted, $\mathbb{P}(\sigma_B)$ denotes the probability that the arriving customers B appear in arrival sequence of permutation σ_B, and $Profit(\sigma_B)$ denotes the total profit obtained by arriving customers B in arrival sequence σ_B. Here, $rev_C(\sigma_B)$ denotes the revenue (including the delivery fees of non-contacted customers) of *all* served customers in the arrival sequence of permutation σ_B, $cost(\sigma_B)$ denotes the cost of inserting *all* feasible customers into the already existing route, given the arrival sequence of permutation σ_B, $pen_C(\sigma_B)$ denotes the penalties of *all* contacted customers that have to be rejected based on the arrival sequence of permutation σ_B, and $opp_C(\sigma_B)$ denotes the opportunity costs of serving *all* inserted contacted customers based on the arrival sequence of permutation σ_B.

Decomposition of the Opportunity Costs o_c

Since proactive customer contacting can shift future demand forward to an earlier date, the effects on future routes and customers constitute an essential component of the proactive customer contacting approach. We summarize all these future effects for a contacted customer $c \in \mathcal{C}$ in the opportunity costs o_c. Because the interdependencies over multiple days are entirely reflected in these opportunity costs, the accurate depiction of these costs is of importance, especially with regard to the performance of the contacting approach. It should be noted that these opportunity costs do not only consist of a pure negative component (i.e., displacement of a future customer) but also include a positive effect (released future resources) that mitigates the magnitude of these costs. In the following, we provide a more detailed decomposition of the two main components reflected in the opportunity costs.

1. *Displacement of a future customer:*
   As proactive customer contacting stimulates existing customers to place an order, this approach leads to a shift in future customers. A customer who would have placed an order in the future (e.g., tomorrow) with probability p is moved to an earlier date. This probability might depend on how many days into the future customers are considered. We assume that the e-grocer determines this future time horizon (e.g., three days) based on historical booking frequencies and that the probability p is identical for all future customers. The opportunity costs must account for this displacement of future revenue r_c and the loss of future delivery charges δ. In particular, we assume that shifting the customer forward (e.g., from tomorrow to today) does not affect his expected order (normal shopping basket) and hence the obtained revenues. Accounting for the effect of shifting future customers is essential, as one would otherwise assume that the revenue obtained by arriving contacted customers is auxiliary revenue, which would not be justifiable from a practical perspective. In the same vein, the effect on future routing costs needs to be addressed. Since future customers and hence the future route are still unknown, we denote the future routing cost estimation by $\bar{c}$ (which can be obtained by considering the marginal costs of serving an additional customer on previous routes). The negative component (displacement of a future customers) of the opportunity costs amounts to $p \cdot (r_c + \delta - \bar{c})$.

2. *Released future resources:*
   On the other hand, shifting a future customer to an earlier time releases future resources that would have been consumed by this customer, both in terms of the physical capacity of the vehicle and the tour length. The flexibility gained by preponing the customer can be exploited in the future. Whether these available resources can be utilized by serving another customer, however, depends on the overall system utilization. We represent the system utilization by a parameter u, which indicates the probability that the released resources created by rescheduling a customer can be filled by another customer. Since this probability is difficult to determine, we evaluate several system utilization settings in our numerical study in Section 4.4.6. If the resources can be exploited, we assume an arriving customer with average revenue $\bar{r}$ and delivery fee δ. Because the future customers and hence the future route are again unknown, we approximate the accompanying future routing costs with $\bar{c}$. Accordingly, the positive component (released future resources) of the opportunity costs is aggregated as $u \cdot (\bar{r} + \delta - \bar{c})$.

Combining the negative and positive effects of serving a contacted arriving customer $c \in \mathcal{C}$, the opportunity costs can be calculated as

$$\begin{aligned} o_c &= p \cdot (r_c + \delta - \bar{c}) - u \cdot (\bar{r} + \delta - \bar{c}) \\ &= pr_c - u\bar{r} + (p - u) \cdot (\delta - \bar{c}). \end{aligned}$$

4.4. Numerical Study

In this section, we investigate the proactive customer contacting approach in a numerical study and evaluate the merit of integrating proactive customer contacting under different settings. To this end, we first introduce different benchmarks in Section 4.4.1 to assess the results of the optimal contacting decision (resulting from (4.2) - (4.3)). In particular, we consider the different contacting strategies *Contact No Customer*, *Contact All Customers* and *Greedy Contacting*. In addition, we propose a procedure for obtaining a sampled ex post optimal solution. After clarifying the considered parameters and the general setting used for the numerical study in Sections 4.4.2 and 4.4.3, we evaluate the base scenario in Section 4.4.4 and compare the performance of the different contacting strategies. To determine what drives the success of proactive customer contacting, we then modify the considered contacting setting. In Section 4.4.5, we evaluate how the benefit of the proactive customer contacting approach depends on the preliminary route utilization by altering the number of customers that can be added to the preliminary route. In Section 4.4.6, we examine whether proactive customer contacting is advisable for e-grocers with a generally low utilization by modifying the overall system utilization. In Section 4.4.7, we amend the efficacy of proactive customer contacting by varying the response probability of contacted customers. Section 4.4.8 summarizes the results by formulating the obtained managerial insights.

4.4.1. Benchmarks

4.4.1.1. Simple Contacting Strategies

To assess the performance of the optimal contacting solution, we provide simple contacting schemes. These contacting strategies are *Contact No Customer*, *Contact All Customers*, and *Greedy Contacting*, which we describe next.

1. *Contact No Customer:*
The simplest contacting strategy is not to contact customers at all. In this extreme case, no customer is cannibalized from the normal booking process. However, there is a high probability that there will be spare capacity at the end of the booking horizon. Using this strategy, the profit can be calculated as

$$\text{Profit} = \sum_{B \subset \mathcal{C}} \tilde{\lambda}^{|B|} \cdot (1-\tilde{\lambda})^{|\mathcal{C} \setminus B|} \sum_{\sigma_B \in Perm(B)} \frac{1}{|B|!} \cdot (rev_{\emptyset}(\sigma_B) - cost(\sigma_B)).$$

2. *Contact All Customers:*
In the opposite extreme, the lever of proactive customer contacting is fully engaged, all customers available for contacting are contacted, and offered free delivery. This leads to a high utilization of the system. However, it also increases the risk of having many dissatisfied contacted customers who cannot place an order because the capacity is already exhausted. For this contacting strategy, the profit can be determined as

$$\begin{aligned} \text{Profit} = \sum_{B \subset \mathcal{C}} \lambda^{|B|} \cdot (1-\lambda)^{|\mathcal{C} \setminus B|} \sum_{\sigma_B \in Perm(B)} \frac{1}{|B|!} \\ \cdot (rev_{\mathcal{C}}(\sigma_B) - cost(\sigma_B) - pen_{\mathcal{C}}(\sigma_B) - opp_{\mathcal{C}}(\sigma_B)). \end{aligned}$$

3. *Greedy Contacting:*
A simple contacting strategy that lies between these two extremes is to contact only the "best" customers. The idea of this contacting scheme is based on two steps: first, ranking the customers, and second, determining the number of customers to contact. In the first step, customers are ranked based on their marginal profit, i.e., the profit generated by the order minus the additional routing costs (based on the preliminary route), which customers would contribute if added to the route. In the second step, it is determined how many customers should be contacted. For this step, the number of customers that can be added to the preliminary route is used as a proxy. Using these two steps, the set of "best" customers C^* to contact can be determined, and the profit can be calculated as

$$\begin{aligned} \text{Profit} = \sum_{B \subset \mathcal{C}} \lambda^{|B \cap C^*|} \cdot (1-\lambda)^{|C^* \setminus B|} \cdot \tilde{\lambda}^{|B \setminus C^*|} \cdot (1-\tilde{\lambda})^{|\mathcal{C} \setminus (B \cup C^*)|} \\ \cdot \sum_{\sigma_B \in Perm(B)} \frac{1}{|B|!} \cdot (rev_{C^*}(\sigma_B) - cost(\sigma_B) - pen_{C^*}(\sigma_B) - opp_{C^*}(\sigma_B)). \end{aligned}$$

4.4.1.2. Sampled Ex Post Optimal Solution

To better contrast the performance of the optimal contacting decision with the different contacting strategies and to likewise determine the potential of contacting in light of what would have been the best contacting decision under optimal contacting conditions in the considered environment, we provide a sampled ex post optimal solution to the problem. To this end, we report the results of first sampling which customers arrive (depending on whether they were contacted or not) and then solving a team orienteering problem with time windows (TOPTW) for these customers. For given customer arrivals, this provides an ex post optimal solution for the contacting procedure that naturally avoids contacted customer rejections. In particular, we use the TOPTW formulation from Vansteenwegen et al. (2009) and adapt it to our specific problem setting. The detailed model formulation and description can be found in the Appendix. The goal of this model is to find a feasible tour (ensuring that all customers already accepted on the preliminary route are served, and the maximum tour length and physical capacity of the vehicle are not exceeded) that maximizes the additional profit, i.e., the revenue of the additionally visited customers (including the delivery fees of the non-contacted customers) minus the opportunity costs of the contacted customers and the additional routing costs compared to the preliminary route. The results of the sampled ex post optimal solution may not provide a tight upper bound on the optimal contacting decision for the following two reasons:

1. *No cannibalization effect:*
   Since the customers are sampled at the beginning of the process, it is assumed that it is already known whether a customer will arrive depending on whether he is contacted. This has a significant impact on the contacting decision, as an arriving customer is never contacted (no cannibalization effect) and the delivery fee is always obtained. In this sampled optimal ex post decision, overbooking and the associated penalties never occur.

2. *Optimal route and optimal customer arrival sequence:*
   Since the TOPTW model formulation determines the route when all customers are present, an underlying optimal arrival sequence of customers is assumed. Compared to the contacting optimization approach, which uses the nearest insertion heuristic to determine the routing cost, this sampled ex post approach determines the optimal route based on the knowledge of the entire arriving customer population.

Although the sampled ex post optimal solution may produce results that are significantly superior to those of the optimal contacting decision because additional knowledge is assumed, this solution is a useful indicator for evaluating the potential of proactive customer contacting and the different contacting strategies under optimal conditions.

4.4.2. Data and Parameters

Before diving into the results of the numerical study, we explain the relevant parameter settings, which are summarized in Table 4.1. Some parameters considered in this numerical study are similar to those in Schwamberger et al. (2022b). As in Schwamberger et al. (2022b), we focus on the area of London, assume an underlying delivery area of $40km$ x $40km$, and locate the depot in the center of the delivery area. The vans that deliver the groceries have a certain physical capacity. Since a van can carry 70 crates (ambient crates, cooled crates and frozen crates) and an average customer order consists of 5 crates, a maximum of 14 customers can be served per tour. Since a delivery shift is 8 hours and the average delivery speed of a vehicle is assumed to be 25 km per hour, the overall tour length is limited to 200 km. In addition to the maximum tour length, a dwell time of 12 minutes is incurred per customer.

In contrast to Schwamberger et al. (2022b), we focus on the additional profit that can be generated by proactive customer contacting in post-crisis times. Thus, we need to include additional cost parameters. We assume that the delivery fee is set to £5 and that the e-grocer incurs costs of £0.4 per km. The revenue of a customer is assumed to be £61 (Statista 2018). Since rejecting contacted customers should be avoided, we set the penalty for rejecting a contacted customer to half of the obtained revenue, i.e., £30.5. The probability that a contacted customer would have arrived in the future is presumed to be 85%. Since we consider an e-grocer that does not have a structural demand deficit, we assume that demand and delivery capacity are fundamentally aligned and suppose that the overall system utilization leads to a probability of 80% that future released resources can be utilized. Based on the average detour, the routing costs of inserting a customer into a future route are set to £10. Furthermore, we presume a response probability of contacted customers of $\lambda = 0.9$ and an arrival probability of non-contacted customers of $\tilde{\lambda} = 0.2$.

Table 4.1.: Parameters

Parameter	**Meaning**	**Value**
$DeliveryArea$	Delivery Area	40 km x 40 km
$MaxCap$	Max. Physical Capacity of Vehicle	14
$MaxLength$	Max. Length of Tour	8 Hours
τ	Service Time at Customer Location	12 Minutes
δ	Delivery Fee	£5
γ	Costs per km	0.4 £/ km
$Revenue$	Revenue per Customer	£61
$Penalty$	Penalty for Rejecting a Contacted Customer	£30.5
p	Probability that a Contacted Customer Would Have Arrived in the Future	0.85
u	Probability that Released Future Resources Can Be Utilized	0.8
$\bar{c}$	Approximated Future Routing Costs	£10
λ	Arrival Probability of Contacted Customers	0.9
$\tilde{\lambda}$	Arrival Probability of Non-Contacted Customers	0.2

4.4.3. Instance Generation & Sampling Process

To ensure that our numerical results are generalizable, we test a variety of predefined routes and different available customers to contact. In total, we consider ten different preliminary routes in each of our numerical studies, each randomly generated. For this purpose, we create uniformly distributed arriving customers (before proactive customer contacting takes place) and plan a route to serve these customers. The time windows associated with these customers have a length of 2 hours. In all instances, unless indicated otherwise, twelve customers are generated. These obtained routes are then used as input for the preliminary route at the moment of contacting. Similar to the preliminary route, we generate random customers that are available for contacting. Here, we generate four potential customers for each instance. This number of customers available for contacting is sufficient, as we will see in the following sections by looking at the number of customers that are contacted. Increasing the number of customers to contact would further complicate the selection process and thus the determination of which customers to optimally contact. We refrain from increasing the number of customers and implicitly assume that these customers reflect the best set of customers available for contacting.

To determine the sampled ex post optimal solution, we again start with the previously described preliminary routes and generate 100 simulations regarding which of the customers that are available for contacting would have arrived (with or without being contacted). Subsequently, in each of these simulations, the TOPTW is solved to optimality, and the average achieved profit is calculated.

4.4.4. Base Case

Since we want to evaluate the benefit of proactive customer contacting, we first discuss the results of the considered base case. To this end, we compare the results of the optimal contacting decision (resulting from (4.2) - (4.3)) with the different contacting strategies and the sampled ex post optimal solution. The results are summarized in Table 4.2 and depict the average number of customers contacted and the average profit achieved by the different contacting methods for all considered routes.

Comparing the results listed in Table 4.2, it becomes apparent that the different contacting strategies perform considerably different. In our considered base case, the worst results are obtained when no customer is contacted, followed by contacting all customers. This is not surprising, as these are the two extreme options.

Table 4.2.: Numerical Results: Base Case

Method	**# Cont. Cust.**	**Profit**
Optimization	2	103.77
All Customers	4	65.39
No Customer	0	50.07
Greedy	2	103.48
Ex Post	1.2	120.93

What is remarkable, however, is how well the contacting strategy *Greedy Contacting* performs. Although the optimal solution does not always coincide with the set of customers identified by this contacting strategy, as can be observed in Table 4.2, this simple contacting strategy leads to very good results. The suboptimality of this approach (albeit not very salient) demonstrates that contacting the marginal best customers is not always optimal. Rather, it can be advantageous to contact customers in bundles. For example, it may be beneficial to contact a set of customers that can collectively be well integrated into the current route. Thus, the optimal contacting decision and the success of proactive customer contacting depends on the interaction of customers available for contacting.

Table 4.2 also reports the results of the sampled ex post optimal solution. These results can be regarded as the best possible outcome achieved by proactive customer contacting under optimal conditions without any cannibalization effect and with an optimal customer arrival sequence, which can be used to characterize the potential of contacting in the base case. Comparing this figure with that obtained by the optimization approach, it can be observed that these two effects (cannibalization & arrival sequence) have a significant impact, as expected. Reviewing the figures of the optimal contacting approach, the other simple contacting strategies and the sampled ex post optimal solution, it becomes obvious that proactive customer contacting provides the potential to exploit available untapped resources.

4.4.5. Impact of Preliminary Route Utilization

The success of proactive customer contacting inevitably depends on the utilization of the preliminary route. In this section, we examine the impact of utilization by modifying the number of customers that can be added to the already existing route. In particular, Table 4.3 presents the results of the different contacting strategies and the sampled ex post optimal solution when the preliminary route provides capacity for one, two and three further customers. The base case is shown in bold.

Table 4.3.: Numerical Results: Different Numbers of Insertable Customers

# Insertable Cust.	Method	# Cont. Cust.	Profit
1	Optimization	1	49.83
	All Customers	4	-21.97
	No Customer	0	38.46
	Greedy	1	49.83
	Ex Post	0.2	62.48
2	**Optimization**	**2**	**103.77**
	All Customers	**4**	**65.39**
	No Customer	**0**	**50.07**
	Greedy	**2**	**103.48**
	Ex Post	**1.2**	**120.93**
3	Optimization	3	154.59
	All Customers	4	145.76
	No Customer	0	51.89
	Greedy	3	154.59
	Ex Post	2.2	175.45

Considering the results in Table 4.3, it becomes apparent that if the preliminary route has spare capacity for one, two or three further customers, one, two or three customers are optimally contacted. Thus, in the considered setting, no overbooking occurs. Comparing the profits, we observe that the different contacting strategies perform considerably different, depending on the number of customers that can be added to the preliminary route. What is particularly salient is that the contacting strategies *Contact All Customers* and *Contact No Customer* yield very different results and depend strongly on the overall problem setting, which determines the superiority of one of the approaches. For example, in the case in which the vehicle can serve only one further customer, the contacting strategy *Contact All Customers* leads to negative results, as too many customers are contacted and many of them have to be rejected, whereas the results of the contacting strategy *Contact No Customer* are close to the results of the optimal solution. The opposite occurs if three further customers can be inserted into the preliminary route. Under these circumstances, the contacting strategy *Contact All Customers* yields nearly optimal results, whereas *Contact No Customer* results in only one-third of the optimal profit. Considering the solutions of the contacting strategy *Greedy Contacting*, we observe that it leads to almost optimal results in all different settings considered here.

In principle, the more customers can be integrated into the preliminary route, the more customers will optimally be contacted and served, leading to a higher profit. Since we observe a significantly increasing difference in profit achieved by optimal contacting compared to contacting no customer as the number of customers that can be included in the preliminary route increases, we infer that the utilization of the preliminary route is a substantial driver of the benefit of proactive customer contacting.

4.4.6. Impact of Overall System Utilization

Because proactive customer contacting provides an incentive for potential future customers to place their orders earlier, it is important to scrutinize the impact of the overall system utilization on the optimal contacting decision. For this purpose, we modify the probability u that reflects how likely it is that future released resources, resulting from a displacement of a customer, can be filled by another customer. Table 4.4 displays the results if u is set to 0.8, 0.5, 0.2, and 0.

We observe that if we assume a high or medium overall system utilization and set $u = 0.8$ or $u = 0.5$, two customers are optimally contacted. If we lower the system utilization and consider $u = 0.2$, one customer is optimally contacted. In the extreme case of very low system utilization ($u = 0$), it is optimal to not contact any customer at all. In this setting, it is not reasonable to offer free delivery to a customer if it is assumed that the released future resources cannot be exploited. Based on the profits outlined in Table 4.4, it becomes clear that the contacting strategy *Greedy Contacting* does not inherently provide good contacting decisions. If, for example, the system utilization is lowered, this contacting strategy is not competitive. In the setting of $u = 0.2$, not contacting any customer yields better results.

Table 4.4.: Numerical Results: Different System Utilization

u	Method	# Cont. Cust.	Profit
	Optimization	**2**	**103.77**
	All Customers	**4**	**65.39**
0.8	**No Customer**	**0**	**50.07**
	Greedy	**2**	**103.48**
	Ex Post	**1.2**	**120.93**
	Optimization	2	76.98
	All Customers	4	31.87
0.5	No Customer	0	50.07
	Greedy	2	76.68
	Ex Post	1.2	100.74
	Optimization	1	54.64
	All Customers	4	-1.66
0.2	No Customer	0	50.07
	Greedy	2	49.89
	Ex Post	1.2	81.15
	Optimization	0	50.07
	All Customers	4	-24.01
0	No Customer	0	50.07
	Greedy	2	32.02
	Ex Post	1.2	65.65

Counterintuitively, the number of customers contacted by the sampled ex post optimal solution is not zero in the extreme scenario of very low system utilization $u = 0$. This is because this method first samples the customer arrival process and only contacts customers if not enough customers materialize. This implies that this method has complete knowledge of which customers will be included in the route. Because there is no detrimental cannibalization effect in this setting, it makes sense to fill a gap in the current route with a hole in a future route (and waive the delivery fee) if the cost of adding this customer to today's route is lower than the anticipated additional cost of adding the customer to a future route minus the delivery fee. This, however, is in strong contrast to the optimal contacting decision, since the knowledge of all arriving customers is not available at the moment of contacting.

Examining the profits achieved by the different contacting strategies in Table 4.4, we can deduce that the benefit of contacting significantly depends on the overall system utilization. In particular, we observe that as the system utilization decreases, the achieved profit of the optimal contacting decision decreases. In the extreme case of very low system utilization, it is optimal to not contact any customer at all, resulting in no additional benefit of the demand management lever of proactive customer contacting.

4.4.7. Impact of the Response Probability of Contacted Customers

It is not certain that contacted customers will place orders. Thus, we investigate the impact of the response probabilities of contacted customers on the results of the proactive customer contacting approach by considering scenarios with high, medium and low response likelihoods. The results are summarized in Table 4.5, displaying the figures for the different response probabilities $\lambda = 0.9$, $\lambda = 0.6$, $\lambda = 0.3$, and $\lambda = 0.2$.

Table 4.5.: Numerical Results: Different Response Probabilities

λ	Method	# Cont. Cust.	Profit
	Optimization	**2**	**103.77**
	All Customers	**4**	**65.39**
0.9	**No Customer**	**0**	**50.07**
	Greedy	**2**	**103.48**
	Ex Post	**1.2**	**120.93**
	Optimization	3	87.36
	All Customers	4	83.92
0.6	No Customer	0	50.07
	Greedy	2	82.24
	Ex Post	1.2	109.46
	Optimization	3.9	60.38
	All Customers	4	60.34
0.3	No Customer	0	50.07
	Greedy	2	56.31
	Ex Post	0.96	69.53
	Optimization	0	50.07
	All Customers	4	43.11
0.2	No Customer	0	50.07
	Greedy	2	46.63
	Ex Post	0	50.71

We observe that as the response probability of contacted customers decreases, the aggressiveness of contacting in terms of the number of customers contacted first increases. This becomes apparent when considering the scenario in which customers are likely to place an order after being contacted, i.e., $\lambda = 0.9$, and comparing it to the scenarios with lower response probabilities, i.e., $\lambda = 0.6$ and $\lambda = 0.3$. In the setting where the response probability is slightly higher than the arrival probability of non-contacted customers ($\lambda = 0.3$), it is optimal to contact all customers in almost all cases. This is intuitive, since more customers need to be contacted to obtain enough arriving customers to place an order. However, when the response probability is no longer significantly different from the arrival probability of non-contacted customers or even coincides with it ($\lambda = 0.2$), the number of contacted customers decreases and it may even be optimal to contact no customer. If it cannot be assumed that offering free deliveries provides an incentive for customers to place an order, delivery should not be offered free of charge.

Considering the profits presented in Table 4.5, the lower the response probability of the contacted customers is, the lower the expected profits from proactive customer contacting, independent of the number of customers contacted. Similarly, the difference from the contacting strategy *Contact No Customer* decreases as the response probability decreases, meaning that the benefit of contacting decreases with a decreasing response probability. In this regard, contacting is only beneficial if it can be assumed that customers will respond to contacting and will be enticed to place an order by receiving free delivery.

4.4.8. Managerial Insights

We conclude the numerical study by formulating the obtained managerial insights. In almost all our conducted experiments, we demonstrate that proactive customer contacting can exploit available untapped resources. In particular, if an e-grocer has already implemented the tool of contacting customers during the COVID-19 pandemic, it is a valuable enhancement to the normal booking process that should be leveraged. However, if an e-grocer has not yet deployed proactive customer contacting, it is necessary to carefully assess whether this is a profitable idea. Although we observed in the numerical experiments that proactive contacting can increase the achieved profit in the magnitude of approximately one additional customer per route in the base case, we likewise saw that the benefit of contacting decreases quickly for various less favorable parameter settings. Putting the additional benefit in relation to the effort required to implement and execute proactive customer contacting, this demand management lever might not be the desired game changer.

When contacting customers is reasonable, it is not always optimal to contact the marginally best customers. Of course, if only one customer can be added to the tour, this is the optimal decision. However, if multiple customers can be included, it might be beneficial to deviate from this approach. For example, it may be beneficial to contact a bundle of customers that can collectively be well integrated into the current route. Therefore, the optimal contacting decision always depends on the interrelationship of customers available for contacting. From the numerical results, it can also be deduced that the different simple contacting strategies lead to completely different results, depending on the individual problem setting. While for some instances the contacting strategy *Greedy Contacting* is superior, for other instances the contacting strategies *Contact No Customer* or *Contact All Customers* are dominant. Thus, none of the simple contacting strategies studied is universally superior to the others.

We also identified certain drivers that significantly impact the benefit of proactive customer contacting. First, we observe that the more customers can be added to the preliminary route, the more customers are optimally contacted and served. In particular, the benefit of proactively contacting customers increases if the preliminary route can serve further customers. This is reasonable, as the capacity and flexibility of the current route can be further exploited, especially when the overall system utilization is high and the generated future released resources can be filled with other customers. In all considered settings, it became obvious that the higher the overall system utilization is, the higher the benefit of proactive customer contacting. We infer that contacting is not always a reasonable opportunity but should only be exploited if e-grocers have overall high system utilization, meaning that the released future resources can be utilized by other customers. Otherwise, a gap in the current route is compensated by a gap in a future route, and no delivery fees are charged. It was also found that the method of contacting changes depending on the response probability of the contacted customers. If the response probability is lowered, more customers are contacted initially to utilize the available resources. However, if the response probability is reduced further and approaches the arrival probability of non-contacted customers, fewer or no customers are contacted, as it is pointless to offer free deliveries if contacting does not prompt customers to place an order. In any case, as the response probability decreases, the benefit of integrating proactive customer contacting diminishes.

In summary, the obtained managerial insights are as follows:

- Proactive customer contacting can exploit available untapped resources but might not be the desired game changer.
- It is not always optimal to target the marginal best customers; instead, the optimal contacting decision depends on the interrelationship of customers available for contacting.
- None of the simple contacting strategies (*Contact No Customer*, *Contact All Customers*, and *Greedy Contacting*) is universally superior to the others.
- The benefit of contacting increases if the preliminary route provides more flexibility.
- The higher the overall system utilization is, the higher the benefit of proactive customer contacting.
- If the response probability of the contacted customer decreases, the number of customers to contact first increases and then decreases. In any case, the lower the response probability of the contacted customers is, the smaller the benefit of proactive customer contacting.

4.5. Conclusion

To conclude, we investigate how the concept of proactive customer contacting that was developed during the COVID-19 pandemic can be leveraged for an e-grocer in post-crisis times. In particular, we propose to integrate the customer contacting at the end of the classical booking process and to use this demand management lever as an emergency tool for situations in which it is anticipated that insufficient inquiries will emerge via the normal booking process. We develop an optimization model to determine whether and which customers to contact and compare the results of the optimal contacting decision with those of several contacting strategies and a sampled ex post optimal solution in an extensive numerical study.

We infer that integrating proactive customer contacting in post-crisis times can achieve an increase in profit but might not be the desired game changer, as it has to be put in perspective with the effort of implementing and executing customer contacting. We observe that targeting the marginal best customers is generally not optimal, as the optimal contacting decision depends on the set of customers available for contacting. In addition, we find that none of the simple contacting approaches is universally superior. Overall, we conclude that the benefit of integrating proactive customer contacting strongly depends on the preliminary route utilization, the overall system utilization and the response probabilities of the contacted customers.

Since the idea of using proactive customer contacting as a demand management lever in the context of e-groceries is still young and this research field is relatively unexplored, we see this work as an initial starting point that lays the foundation for further research. For example, it might be interesting to investigate the problem of proactively contacting customers in a multi-period setting. This would provide a better estimate of the opportunity costs of the inserted contacted customers but would also significantly complicate the whole analysis, as the next day's routing issues would need to be considered as well. We leave this extension to future research. From a practical perspective, it would also be helpful to observe whether this proactive contacting approach and the incentive of free deliveries are likely to be accepted by customers. This could be validated with real data. In this regard, future research could integrate more sophisticated customer behavior by incorporating dependencies on the offered discounts into the model. In our numerical study, we assumed that we can modify individual parameters to evaluate parameter-specific effects. In reality, some of these modified parameters are inherently interrelated.

Future research could investigate the correlations of the considered parameters in more detail, allowing us to incorporate these implicit interdependencies into our numerical study.

Appendix A.

Sampled Ex Post Optimal Solution

The sampled ex post optimal solution is obtained by using the sampling process described in Section 4.4.1.2 and solving a modified version of the TOPTW formulation from Vansteenwegen et al. (2009). In this section, we elaborate on the model formulation used. We denote the set of already accepted customers from the preliminary route as C_{pre}, the sampled set of customers that arrive without being contacted as C_{nocont} and the sampled set of arriving contacted customers as C_{cont}. The union of all of these sets is denoted by $C_{all} = C_{pre} \cup C_{nocont} \cup C_{cont}$. If the depot is added to this set, we use the notation C_{all}^{Depot}. By L_{pre}, we denote the length of the preliminary route, and d_{ij} denotes the distance between locations $i \in C_{all}^{Depot}$ and $j \in C_{all}^{Depot}$. All other parameters are the same as in the problem definition (Section 4.3).

We use the following decision variables:

- $x_{ij} \in \{0, 1\}$: a visit to location $i \in C_{all}^{Depot}$ is followed by a visit to location $j \in C_{all}^{Depot}$
- $y_i \in \{0, 1\}$: customer $i \in C_{all}$ is visited on the route
- s_i: start of service for customer $i \in C_{all}$

With this notation, we formulate the TOPTW as

$$\max \sum_{i \in C_{nocont}} (r_i + \delta) \cdot y_i + \sum_{i \in C_{cont}} (r_i - o_i) y_i - \gamma \cdot (\sum_{i \in C_{all}^{Depot}} \sum_{j \in C_{all}^{Depot}} d_{ij} x_{ij} - L_{pre}) \tag{A.1}$$

$$\sum_{j \in C_{all}} x_{Depotj} = 1 \tag{A.2}$$

$$\sum_{i \in C_{all}} x_{iDepot} = 1 \tag{A.3}$$

$$\sum_{i \in C_{all}} y_i \leq MaxCap \tag{A.4}$$

$$\sum_{i \in C_{all}^{Depot}} x_{ij} = y_j \qquad \forall\, j \in C_{all} \tag{A.5}$$

$$\sum_{j \in C_{all}^{Depot}} x_{ij} = y_i \qquad \forall\, i \in C_{all} \tag{A.6}$$

$$\sum_{i \in C_{all}} \tau y_i + \sum_{i \in C_{all}^{Depot}} \sum_{j \in C_{all}^{Depot}} d_{ij} x_{ij} \leq MaxLength \tag{A.7}$$

$$y_i = 1 \qquad \forall\, i \in C_{pre} \tag{A.8}$$

$$s_i + \tau + d_{ij} - s_j \leq M(1 - x_{ij}) \qquad \forall\, i, j \in C_{all}^{Depot} \tag{A.9}$$

$$a_i \leq s_i \qquad \forall\, i \in C_{all} \tag{A.10}$$

$$s_i \leq b_i \qquad \forall\, i \in C_{all} \tag{A.11}$$

$$x_{ij} \in \{0, 1\} \qquad \forall\, i, j \in C_{all}^{Depot} \tag{A.12}$$

$$y_i \in \{0, 1\} \qquad \forall i \in C_{all}. \tag{A.13}$$

The objective is to maximize the additional profit consisting of the collected revenues of the non-contacted customers (including delivery fees) and the revenues of the contacted customers minus the opportunity costs of the served contacted customers and minus the additional routing costs compared to those of the preliminary route. Note that the term L_{pre} may be omitted for optimization purposes since it is constant and determined by the preliminary route. Constraints (A.2) and (A.3) ensure that the tour starts and ends at the depot. Constraint (A.4) guarantees that the maximum physical capacity of the vehicle is not exceeded. Constraints (A.5) and (A.6) link the decision variables x and y. Constraint (A.7) ensures that the maximum route length is not exceeded, whereas constraints (A.8) guarantee that all already accepted customers from the preliminary route are served. Constraints (A.9) link the variables x and s. Note that these constraints require a sufficiently large M. Conditions (A.10) - (A.13) define the domains of the decision variables.

Bibliography

Agatz, N., Campbell, A., Fleischmann, M., and Savelsbergh, M. (2011). Time Slot Management in Attended Home Delivery. *Transportation Science*, 45(3):435–449.

Agatz, N., Campbell, A. M., Fleischmann, M., and Savelsbergh, M. (2008a). Challenges and Opportunities in Attended Home Delivery. In: Golden, B. L., Raghavan, S., and Wasil, E., editors, The Vehicle Routing Problem: Latest Advances and New Challenges. *Operations Research/Computer Science Interfaces, Springer*, 43:379–396.

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

Agatz, N., Fleischmann, M., and van Nunen, J. (2008b). E-fulfillment and multi-channel distribution - A review. *European Journal of Operational Research*, 187(2):339–356.

Aouad, A., Farias, V. F., and Levi, R. (2020). Assortment Optimization Under Consider-Then-Choose Choice Models. *Management Science*, 67(6):3368–3386.

Aouad, A., Feldman, J., Segev, D., and Zhang, D. (2019). The Click-Based MNL Model: A Novel Framework for Modeling Click Data in Assortment Optimization. *SSRN Electronic Journal*.

Berger, P. and Nasr, N. (1998). Customer Lifetime Value: Marketing Models and Applications. *Journal of Interactive Marketing*, 12(1):17–30.

Bertsimas, D. and Mišić, V. V. (2019). Exact First-Choice Product Line Optimization. *Operations Research*, 67(3):651–670.

Bose, I. and Chen, X. (2009). Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research*, 195(1):1–16.

Campbell, A. M. and Savelsbergh, M. W. P. (2005). Decision Support for Consumer Direct Grocery Initiatives. *Transportation Science*, 39(3):313–327.

Choi, T. M. (2020). Innovative "Bring-Service-Near-Your-Home" operations under Corona-Virus (COVID-19/SARS-CoV-2) outbreak: Can logistics become the Messiah? *Transportation Research Part E: Logistics and Transportation Review*, 140:1–17.

Cleophas, C. and Ehmke, J. F. (2014). When are deliveries profitable? *Business and Information Systems Engineering*, 6(3):153–163.

CNA (2020). Delivery slots run dry as more turn to online grocery shopping amid COVID-19 concerns. Channel News Asia. Last accessed: February 22, 2023, `https://www.channelnewsasia.com/singapore/covid-19-coronavirus-groceries-supermarkets-delivery-slots-full-1326196`.

Daganzo, C. F. (1987). Modeling Distribution Problems with Time Windows: Part I. *Transportation Science*, 21(3):171–179.

Department of Finance, UK (2020). Covid-19 Priority Online Shopping Slots. Department of Finance, UK. Last accessed: February 22, 2023, `https://www.finance-ni.gov.uk/articles/covid-19-priority-online-shopping-slots`.

Ehmke, J. F. and Campbell, A. M. (2014). Customer acceptance mechanisms for home deliveries in metropolitan areas. *European Journal of Operational Research*, 233(1):193–207.

Farias, V. F., Jagabathula, S., and Shah, D. (2013). A nonparametric approach to modeling choice with limited data. *Management Science*, 59(2):305–322.

Feldman, J., Paul, A., and Topaloglu, H. (2019). Technical Note - Assortment Optimization with Small Consideration Sets. *Operations Research*, 67(5):1283 – 1299.

Fleckenstein, D., Klein, R., and Steinhardt, C. (2023). Recent advances in integrating demand management and vehicle routing: A methodological review. *European Journal of Operational Research*, 306(2):499–518.

Forbes (2022). How Grocery Retailers Can Navigate E-Groceries In An Inflationary World. Forbes. Last accessed: February 22, 2023, `https://www.forbes.com/sites/forbesbusinesscouncil/2022/10/17/how-grocery-retailers-can-navigate-e-groceries-in-an-inflationary-world/?sh=28ab3bab5f7f`.

Gallego, G. and Li, A. (2017). Attention, Consideration then Selection Choice Model. *SSRN Electronic Journal*.

Gilbride, T. and Allenby, G. (2004). A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. *Marketing Science*, 23(3):391–406.

Grosfeld-Nir, A. and Gerchak, Y. (2004). Multiple Lotsizing in Production to Order with Random Yields: Review of Recent Advances. *Annals of Operations Research*, 126:43–69.

Hauser, J. R. (2014). Consideration-set heuristics. *Journal of Business Research*, 67(8):1688–1699.

Jagabathula, S., Mitrofanov, D., and Vulcano, G. (2019). Inferring Consideration Sets from Sales Transaction Data. *SSRN Electronic Journal*.

Jagabathula, S. and Rusmevichientong, P. (2017). A nonparametric joint assortment and price choice model. *Management Science*, 63(9):3128–3145.

Keskin, M., Branke, J., Deineko, V., and Strauss, A. (2023). Dynamic Multi-Period Vehicle Routing with Touting. *European Journal of Operational Research*.

Klein, R., Koch, S., Steinhardt, C., and Strauss, A. (2020). A review of revenue management: Recent generalizations and advances in industry applications. *European Journal of Operational Research*, 284(2):397–412.

Klein, R., Mackert, J., Neugebauer, M., and Steinhardt, C. (2018). A model-based approximation of opportunity cost for dynamic pricing in attended home delivery. *OR Spectrum*, 40(4):969–996.

Klein, R., Neugebauer, M., Ratkovitch, D., and Steinhardt, C. (2017). Differentiated Time Slot Pricing Under Routing Considerations in Attended Home Delivery. *Transportation Science*, 53(1):236–255.

Kleinrock, L. (2010). An Early History of the Internet. *IEEE Communications Magazine*, 48:26–36.

Landrigan, M. (2005). Customer Relationship Management: Concepts and Tools. *Journal of Consumer Marketing*, 22(4):237–238.

Lang, M. A., Cleophas, C., and Ehmke, J. F. (2021a). Anticipative Dynamic Slotting for Attended Home Deliveries. *Operations Research Forum*, 2(4):70.

Lang, M. A., Cleophas, C., and Ehmke, J. F. (2021b). Multi-criteria decision making in dynamic slotting for attended home deliveries. *Omega*, 102.

L.E.K. (2019). Why Now Is the Time for Grocers and CPGs to Collect on "Click and Collect". Technical report, L.E.K. Consulting / Executive Insight.

Liu, Q. and Arora, N. (2011). Efficient Choice Designs for a Consider-Then-Choose Model. *Marketing Science*, 30(2):321–338.

Mackert, J. (2019). Choice-based dynamic time slot management in attended home delivery. *Computers and Industrial Engineering*, 129:333–345.

Manzini, P. and Mariotti, M. (2014). Stochastic Choice and Consideration Sets. *Econometrica*, 82(3):1153–1176.

McKinsey (2020). How European shoppers will buy groceries in the next normal. McKinsey & Company. Last accessed: February 22, 2023, `https://www.mckinsey.com/industries/retail/our-insights/how-european-shoppers-will-buy-groceries-in-the-next-normal`.

Meissner, J., Strauss, A., and Talluri, K. (2013). An Enhanced Concave Program Relaxation for Choice Network Revenue Management. *Production and Operations Management*, 22(1):71–87.

Mintel (2021). Online Grocery Retailing - UK - March 2021. Technical report, Mintel.

Mkansi, M., Eresia-Eka, C., and Emmanuel-Ebikake, O. (2018). E-Grocery challenges and remedies: Global market leaders perspective. *Cogent Business & Management*, 5(1):1–28.

Oetken, C. N., Yildiz, B., Arsland, O., and Laporte, G. (2022). Making Opportunity Sales in Attended Home Delivery. *SSRN Electronic Journal*.

Paul, A., Topaloglu, H., and Feldman, J. (2016). Assortment Optimization for Choosy Customers. *Working Paper*.

Paul, J., Agatz, N., and Savelsbergh, M. (2019). Optimizing Omni-Channel Fulfillment with Store Transfers. *Transportation Research Part B: Methodological*, 129:381–396.

Pavlik, J. A., Ludden, I. G., Jacobson, S. H., and Sewell, E. C. (2021). Airplane seating assignment problem. *Service Science*, 13(1):1–18.

Reinartz, W. and Venkatesan, R. (2008). Decision Models for Customer Relationship Management (CRM). *International Series in Operations Research & Management Science*, 121:291–326.

Rewe (2020). Klinikpersonal bekommt Lebensmittel bevorzugt und gebührenfrei vom REWE Lieferservice zugestellt. Rewe. Last accessed: February 22, 2023, `https://www.rewe-group.com/de/presse-und-medien/newsroom/pressemitteilungen/klinikpersonal-bekommt-lebensmittel-bevorzugt-und-gebuehrenfrei-vom-rewe-lieferservice-zugestellt/`.

Schwamberger, J. (2023). Getting the Best out of Online Deliveries - Proactive Customer Contacting in E-Fulfillment. *SSRN Electronic Journal*.

Schwamberger, J., Fleischmann, M., and Strauss, A. (2022a). Dataset: Feeding the Nation - Dynamic Customer Contacting for E-Fulfillment in Times of Crisis. Mendeley Data. Last accessed: February 22, 2023, `https://data.mendeley.com/datasets/ptxy5zp9ss/3`.

Schwamberger, J., Fleischmann, M., and Strauss, A. (2022b). Feeding the Nation - Dynamic Customer Contacting for E-Fulfillment in Times of Crisis. *Service Science*, 15(1):22–40.

Schwamberger, J., Fleischmann, M., and Strauss, A. (2023). Tractable Time Slot Assortment Optimization in Attended Home Delivery under Consider-Then-Choose Customer Choice. *SSRN Electronic Journal*.

Statista (2018). Average weekly household expenditure on food in the United Kingdom (UK) in 2017/2018, by age of household reference person (in GBP). Technical report, Statista.

Statista (2022). Marktvolumen des E-Commerce in den Jahren 2006 bis 2021 sowie eine Prognose für 2022. Technical report, Statista.

Strauss, A., Klein, R., and Steinhardt, C. (2018). A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research*, 271(2):375–387.

Strauss, A. K. and Talluri, K. (2017). Tractable Consideration Set Structures for Assortment Optimization and Network Revenue Management. *Production and Operations Management*, 26(7):1359–1368.

Talluri, K. (2010). A Randomized Concave Programming Method for Choice Network Revenue Management. *Working Paper*.

Tang, H. and Miller-Hooks, E. (2005). A TABU search heuristic for the team orienteering problem. *Computers and Operations Research*, 32(6):1379–1407.

Telegraph (2020). 'I can't get a supermarket home delivery slot - what should I do?' Coronavirus Q&A. The Telegraph. Last accessed: February 22, 2023, `https://www.telegraph.co.uk/news/0/home-food-delivery-supermarket-opening-times-rules-tesco-waitrose/`.

Van Ryzin, G. and Vulcano, G. (2015). A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science*, 61(2):281–300.

Van Ryzin, G. and Vulcano, G. (2017). Technical Note—An Expectation-Maximization Method to Estimate a Rank-Based Choice Model of Demand. *Operations Research*, 65(2):396–407.

Vansteenwegen, P., Souffriau, W., and Oudheusden, D. V. (2011). The orienteering problem: A survey. *European Journal of Operational Research*, 209(1):1–10.

Vansteenwegen, P., Souffriau, W., Vanden Berghe, G., and Van Oudheusden, D. (2009). Iterated local search for the team orienteering problem with time windows. *Computers and Operations Research*, 36:3281–3290.

Vox (2020). How to get groceries when delivery services are slammed. Vox. Last accessed: February 22, 2023, `https://www.vox.com/the-goods/2020/4/8/21213919/grocery-delivery-slammed-coronavirus-freshdirect-instacart`.

Wassmuth, K., Köhler, C., Agatz, N., and Fleischmann, M. (2023). Demand management for attended home delivery - A literature review. *European Journal of Operational Research*.

Yang, X. and Strauss, A. (2017). An approximate dynamic programming approach to attended home delivery management. *European Journal of Operational Research*, 263(3):935–945.

Yang, X., Strauss, A., Currie, C., and Eglese, R. (2016). Choice-Based Demand Management and Vehicle Routing in E-Fulfillment. *Transportation Science*, 50(2):473–488.

Yano, C. A. and Lee, H. L. (1995). Lot Sizing with Random Yields: A Review. *Operations Research*, 43(2):311–334.

Yildiz, B. and Savelsbergh, M. (2020). Pricing for delivery time flexibility. *Transportation Research Part B: Methodological*, 133:230–256.

Yorkshire Post (2020). Vulnerable customers: Sainsbury's, Tesco, Asda and other supermarkets are prioritising vulnerable people for delivery slots - this is how to register. Yorkshire Post. Last accessed: February 22, 2023, https://www.yorkshirepost.co.uk/lifestyle/food-and-drink/vulnerable-customers-sainsburys-tesco-asda-and-other-supermarkets-are-prioritising-vulnerable-people-delivery-slots-how-register-2533406.

Curriculum Vitae

Jonas Schwamberger

Professional Experience:

2018 - 2023	Teaching & Research Assistant Chair of Supply Chain Management, University of Mannheim, Germany

Education:

2018 - 2023	Doctoral Studies in Business Administration (Dr. rer. pol.) University of Mannheim, Germany
2016 - 2018	Mathematics (M.Sc.) RWTH Aachen University, Germany
2017 - 2018	Mathematics (Exchange Semester) Beijing Institute of Technology, China
2013 - 2016	Mathematics (B.Sc.) RWTH Aachen University, Germany
2005 - 2013	Abitur Gymnasium Petrinum Brilon, Germany