Dynamic Pricing of Flexible Time Slots for Attended Home Delivery

Arne Strauss[a], Nalan Gülpınar[b], Yijun Zheng[b]

[a]*Mercator Endowed Chair of Sustainable Urban Transport, WHU – Otto Beisheim School of Management, 56179 Vallendar, Germany*
[b]*Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom*

Abstract

In e-commerce, customers are usually offered a menu of home delivery time windows of which they need to select exactly one, even though at least some customers may be more flexible. To exploit the flexibility of such customers, we propose to introduce flexible delivery time slots, defined as any combination of such regular time windows (not necessarily adjacent). In selecting a flexible time slot (out of a set of windows that form the flexible product), the customer agrees to be informed only shortly prior to the dispatching of the delivery vehicle in which regular time window the goods will arrive. In return for providing this flexibility, the company may offer the customer a reduced delivery charge and/or highlight the environmental benefits. Our framework also can accommodate customized flexible slots where customers can self-select a set of regular slots in which a delivery may take place.

The vehicle routing problem (VRP) in the presence of flexible time slots bookings corresponds to a VRP with multiple time windows. We build on literature on demand management and vehicle routing for attended home delivery, as well as on flexible products. These two concepts have not yet been combined, and indeed the results from the flexible products literature do not carry over directly because future expected vehicle routing implications need to be taken into account. The main methodological contribution is the development of a tractable linear programming formulation that links demand management decisions and routing cost implications, whilst accounting for customer choice behavior. The output of this linear program provides information on the (approximate) opportunity cost associated with specific orders and informs a tractable dynamic pricing policy for regular and flexible slots. Numerical experiments, based on realistically-sized scenarios, indicate that expected profit may increase significantly depending on demand intensity when adding flexible slots rather than using only regular slots.

Keywords: Revenue management, e-commerce, Flexible product, Dynamic programming

1. Introduction

Globally, business to consumer e-commerce is growing strongly. According to the Ecommerce-Foundation (2016), global growth in turnover has been around 17.5% in 2016. Online grocery retailing,

---

Email address: `arne.strauss@whu.edu` (Arne Strauss)
Corresponding author

in particular, is growing at a similar rate: in the United Kingdom (UK), sales exhibit strong annual growth rates of 14.7% in 2016 with similar rates forecast over the next years by Mintel (2017). Attended home delivery is commonly offered in e-commerce when the ordered items are either bulky (like furniture), or in need of refrigeration (groceries), or for other reasons need to be handed over to the customer in person. One commonly differentiates between same-day and next-day home deliveries; in this paper, we focus on the next-day deliveries where vehicles are being dispatched after all orders have been collected.

In this domain, there is much competitive pressure over the quality of the delivery service which depends on various factors such as length of the delivery time window, fit of the available slots with customers' schedules, reliability of delivery within the promised slot and others. To address the desire for narrow (and thus more convenient) delivery time windows, most UK grocery retailers (such as Tesco, Sainsbury's, Morrisons, Ocado, Waitrose) are now offering one-hour time slots, as opposed to the longer time windows offered in the past. Such narrow delivery time windows lead to high fulfilment costs because there is little flexibility to make vehicle routes more efficient. At the same time, the cost cannot be easily passed on to the customers because they tend to be very sensitive to delivery charges. This is also reflected by market research conducted by Mintel (2017) showing that current online grocery shoppers, lapsed shoppers and non-users would be most encouraged to buy more online if delivery prices were lower. In other words, delivery charges are a major deterrent from online shopping. The combination of high delivery cost and limited capability to recoup the cost via delivery charges (or via increased product prices) indicates that there is a lot of pressure on making delivery services as efficient as possible; but the high sensitivity to delivery prices also means that delivery prices can be used to influence customers' delivery time slot choice behavior.

Despite the competitive pressure over offering narrow delivery time slots, not all customers actually require them to be so narrow. For example, pensioners, students or people with childcare obligations may be fairly flexible. A significant proportion of people working from home (13.6% of people in employment were working from home in the UK between January 2017 and March 2017 according to the Office for National Statistics (2017)) further underlines this point. Some may be willing to accept uncertainty over the exact delivery time slot within a given set of potential time slots in return for an incentive. This could be a non-monetary incentive by communicating to the customer that more flexible delivery generally entails environmental benefits; and/or the uncertainty might be compensated with a reduced delivery charge. This has recently been exploited by the UK's largest retailer Tesco in that they offer so-called 'Flexi Saver Slots' alongside their regular one-hour slots. Flexi Saver Slots are four hours in length, and the customer is notified on the day of delivery of a one-hour slot (within the booked four-hours window) in which the delivery will be made. The customer pays less in delivery charges in return for giving the retailer more flexibility in their fulfilment operations. This may allow the retailer to accept more orders and/or to fulfil them more efficiently. In this paper, we generalize

this concept further by defining a flexible time slot as any fixed combination of regular delivery time slots, so they do not necessarily need to be adjacent. Moreover, we propose a new version of these flexible slots that we call "customizable flexible slot" that can be managed with our methodology. The idea is to allow customers to pick a combination of a fixed number of available standard slots to create their own flexible slot for a fixed delivery charge discount. For example, this could take the form of "pick any three standard one-hour slots for a delivery charge of only £X". As for any flexible slot, delivery will be carried out in one of the indicated slots. Intuitively, this should be more attractive to customers than only offering wider time slots as flexible slots since they can tailor the slots to their own schedule. We stress that the concept of offering discounts in exchange for flexibility is not the new contribution here (this has been used for a while already by various shippers, such as peapod.com and Tesco; although we are not aware of existing industry applications of a customizable flexible slot). Instead, the challenge lies in quantifying the savings potential of a flexible slot as well as in dynamically pricing these slots under a model that incorporates customer choice behavior. When making a decision on how to price a flexible slot, we need to take into account how much we may be able to save in the routing due to this flexibility, which is difficult to assess because we do not have full information on all orders at the time of making this pricing decision.

In this paper, we study the dynamic pricing problem faced by a firm offering regular and flexible delivery time slots for attended home delivery. We assume that delivery requests for a specific day arrive randomly over a fixed time horizon prior to the delivery day so that the delivery operation takes place after all orders have been received. A delivery request represents a customer who has logged into their account on the retailer's website (hence we know their location) and how has clicked on a link requesting to be shown the available delivery options. The request may come after the shopping basket has been filled such that we also know the value (and estimated size in terms of delivery totes) of the order already; but it might also come beforehand, in which case one can estimate the value and size of the order e.g. by using the mean order value of this customer's previous purchases. We define the value of an order as the profit made on the product sales before delivery costs and excluding potential revenue from a delivery charge.

The decision problem of the retailer consists of (i) evaluating which time windows can feasibly be offered, and (ii) deciding which delivery prices to display for all feasible slots. We assume that all feasible slots are offered so as to increase customer satisfaction, but this assumption can easily be relaxed in our model without structural changes (another dummy price point is needed that drives demand to zero). Delivery charges are assumed to be chosen from a finite set of price points, in line with common industry practice. In response to the firm's decision, the customer chooses a slot or decides to leave without purchase according to a discrete choice model that reflects the set of available options and prices.

After all orders have been received, the firm needs to solve a capacitated vehicle routing problem

with multiple time windows for its fleet of delivery vehicles. Note that 'multiple time windows' refers to having some customers with multiple time windows within which the delivery may take place (namely those customers with flexible slots). The objective is to maximize the profit after delivery cost by dynamic delivery slot pricing and routing of the delivery vehicles to serve the final set of orders. The profit after delivery cost, also called total profit, measures the total revenue of order profit and delivery charges subtracting delivery costs.

Our main contribution is a new approach of how to estimate the opportunity cost associated with accepting a given order in the different delivery options. This opportunity cost reflects the implications on routing costs and potentially displaced profits from future orders in case constraints on the van capacity and/or driving time are binding. The estimation of opportunity cost is very difficult because the calculation of the final delivery cost is challenging even if we already know the final set of orders (which, however, we do not). To tackle this challenge, we propose a novel linear program (LP) that accounts heuristically for both delivery costs and future expected order revenue. This LP is solved offline and should be re-optimized throughout the booking horizon so as to provide updated estimates of the opportunity cost. Furthermore, pricing decisions need to be made in a very short time interval. To that end, we propose an online pricing approach that exploits the features of the choice model and the constraint structure and, thereby, equivalently reduces the nonlinear optimization problem to a small LP that can be solved very quickly. The process of how the offline and online optimization approaches interact over time is outlined in Figure 1. We evaluate the proposed approach in a realistically-sized simulation study and our results show that the concept of flexible delivery slots can significantly improve profitability. Our approach allows for a very general design of flexible slots, meaning that such slots could consist of any combination of standard slots regardless of whether they are adjacent or not. This has the advantage that it allows customizable flexible slots of the type "pick any Y standard slots for a reduced delivery charge of £X".

Figure 1: Solution method
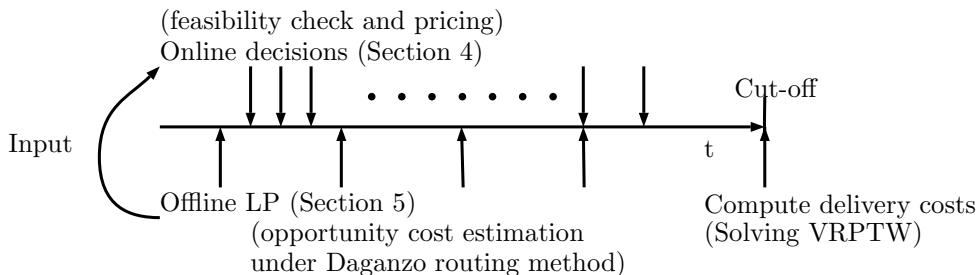
The paper is organized as follows: in Section 2, we review the literature on demand management in the context of attended home deliveries. In Section 3, we define the problem as a Markov decision process and present an intractable dynamic programming formulation that is useful to motivate approximate solution methods. In Section 4, we formulate the pricing policy with flexible slots. In

Section 5, we develop our LP approach to opportunity cost approximation. Section 6 contains the computational results and we draw conclusions and implications for managers in Section 7.

2. Literature Review

We focus on the growing literature on combining demand management with delivery slot booking for attended home deliveries. Demand management in our context is to be understood as optimization of actions that have a direct influence on demand, specifically pricing, deciding on incentives or on the availability of certain offerings. For the reader who is interested in a broader context of e-fulfillment, we refer to the review of Agatz et al. (2008) covering the e-fulfillment literature from an operational research perspective, Agatz et al. (2013) who discuss and classify the revenue management applications in e-fulfillment, and Hübner et al. (2016) who review the recent qualitative fulfillment and distribution literature. We classify the most related research papers developing static and dynamic policies for the attended home delivery systems into two groups in terms of different modelling and solution characteristics as presented in Table 1.

Table 1: Classification of related research papers in attended home delivery

Article	Decision	Revenue Anticipation	Cost Estimation	Choice Model	Slot
Static Policy					
Campbell and Savelsbergh (2005)	availability	-	insertion heuristic	fixed probability	standard
Campbell and Savelsbergh (2006)	pricing	-	insertion heuristic	fixed probability	standard
Cleophas and Ehmke (2014)	availability	simulation	insertion heuristic	-	standard
Klein et al. (2017)	pricing	approx. dynamic program	MIP	nonparametric	standard
Dynamic Policy					
Asdemir et al. (2009)	pricing	dynamic program	-	MNL	standard
Agatz et al. (2011)	availability	-	continuous approximation	-	standard
Ehmke and Campbell (2014)	availability	simulation	insertion heuristic	-	standard
Yang et al. (2016)	pricing	-	insertion heuristic	MNL	standard
Yang and Strauss (2017)	pricing	approx. dynamic program	continuous approximation	MNL	standard
Köhler et al. (2019)	availability	simulation	insertion heuristic	-	short & long
Our paper	pricing	linear program	continuous approximation	MNL	standard & flexible

The first paper to consider aspects of both demand management and delivery operations is Campbell and Savelsbergh (2005) who investigate a dynamic routing and scheduling problem of a grocery vendor who needs to decide which deliveries to accept or reject, and in which time slot to deliver the accepted orders. All customers have a certain time slot profile; this contains all slots that they are willing to accept. If the grocer accepts the order, the company assigns one of these slots to the order. In this first paper, Campbell and Savelsbergh (2005) represent demand as an arrival process that is not affected by the firm's decisions. In their subsequent work, Campbell and Savelsbergh (2006) use relatively simple customer behavior model to include effect of incentives (such as delivery charges) on the probability that a particular time slot is being chosen. The objective is to influence delivery time slot choices to minimize delivery costs, whereas in our work we focus on maximizing expected profit.

A more realistic customer choice model was employed by Asdemir et al. (2009), namely the multinomial logit (MNL). They consider a dynamic time slot pricing approach similar to our paper, but propose dynamic programming (DP) as a solution method with fixed delivery costs rather than our LP-based approach. The DP is formulated at the level of a delivery region (such as a postcode sector) under the assumption that the delivery capacity in each time slot for this region is fixed and known a priori. Practical application of this approach may be challenging when there are many delivery time slots because the DP's state space grows exponentially with the number of slots. In our approach, we also make use of the MNL choice model, but propose a new way of including dynamic delivery cost estimates into a LP model which allows us to solve it for realistically-scaled problem instances.

In contrast to this work on dynamic pricing in attended home delivery, Agatz et al. (2011) focus on the problem of which delivery time slots to offer in which geographic delivery area so as to reduce delivery costs whilst meeting service requirements. They do not consider customer choice behavior, whereas our focus in this paper is on pricing to influence customers' time slot choices. However, there are some common elements in that they also use the work of Daganzo (1987) to obtain a continuous delivery cost approximation.

Another work that stresses the routing and scheduling aspects (as opposed to demand management) in the attended home delivery context is Ehmke and Campbell (2014). Their objective is to maximize the number of requests accepted for delivery, subject to retaining feasible tours. The company makes decisions on accepting or rejecting delivery slot bookings, and the customers' slot choices are assumed to be independent of these controls. Ehmke and Campbell (2014) investigate a simple version of a customizable slot where customers are allowed to propose a preferred time slot and, if it gets rejected, they may propose an alternative slot. However, there is no choice model used in their work to model how customers choose their time slots, nor a methodology on how one should price such constructs. Our setting allows for arbitrary combinations of slots to form customizable slots. In our work, we aim to maximize total expected profit by deciding on prices for regular and flexible delivery slots which influence customers' slot choices. Cleophas and Ehmke (2014) likewise consider accept/reject decisions along with capacity reservations for certain delivery areas and time windows where particularly valuable demand is being expected. Visser and Savelsbergh (2019) decide slot availability based on priori delivery routes constructed under the consideration of future revenue.

Yang et al. (2016) is more closely related to our work in that the authors consider a dynamic pricing problem for delivery slots under the MNL choice model. Using real data, they estimate the choice model and find that demand is very sensitive to delivery prices and slot availability. In their pricing policy, they only rely on opportunity cost estimates based on marginal routing costs (derived by insertion heuristics). They do not consider the effect of future lost revenues due to displaced orders in the opportunity cost estimate. This is addressed by Yang and Strauss (2017) who use an approximate dynamic programming approach to incorporate both future revenue and routing cost effects in the

opportunity cost. Likewise, Koch and Klein (2017) employ approximate dynamic programming to approximate opportunity cost including revenue and cost effects. However, their paper is centered around the idea of quantifying the free delivery time within each time slot for a given route plan. These so-called time budgets are then used to construct value function approximations.

The estimation of routing cost is a major challenge in demand management for attended home delivery. Bühler et al. (2016) discuss various linear mixed-integer programs that approximate the delivery costs for a fixed pool of route candidates. Klein et al. (2017) combine such a linear mixed-integer program (MIP) with the dynamic pricing model of Yang et al. (2016) so as to anticipate future demand. However, the MIP involves a very large number of decision variables for real-life scaled problems, thus making it challenging to solve.

The work of Köhler et al. (2019) is conceptually related in as far as they investigate how to dynamically control the offering of long and/or short delivery time windows to customers in an attended home delivery context. However, here the slots are not 'flexible slots' in the sense that we propose in this paper; instead, their term 'flexible time window management' refers to deciding dynamically which short and/or long slots shall be made available to a given customer (so some customers may be shown only long time windows, other only short ones, and yet others a mix of both). They assume fixed delivery fees for all deliveries regardless of time window length or time of day whereas we use dynamic pricing that reflects both customer preferences and opportunity costs associated with having a customer book any given slot.

In the remainder of this section, we review the literature on flexible products. A flexible product was first proposed by Gallego and Phillips (2004). They define it as a product that can be provided in one of a small number of modes. Customers are aware of this set of modes at the point of purchase, but only receive confirmation of the actual mode at a pre-defined time after purchase (usually shortly prior to product consumption). The product is usually a service such as a flight; in this case, potential modes could correspond to different flights between the same origin and destination but at different departure times. Indeed, Gallego and Phillips (2004) study the problem in the airline context and propose a booking limit control policy for the flexible ticket under a static setting with two time periods and two alternative flights. Gallego et al. (2004) extend this concept to networks, and also consider customer choice modeling. They introduce a deterministic linear program that can approximate the optimal objective of the stochastic optimization problem. Petrick et al. (2010b) likewise propose a deterministic linear program, but focus on independent demand only. They explicitly incorporate the capacity requirements of requests for flexible products that have been previously accepted and thereby allow them to be rearranged.

Among these deterministic linear programming approaches, Petrick et al. (2010a) investigate how they should be used over time to obtain dynamic control mechanisms under independent demand. Gönsch et al. (2014) pursue this further and find that the deterministic linear programming approx-

imation fails to capture the revenue generated from delaying resource allocation by using flexible products. They propose to use the opportunity cost to obtain a dynamic booking limit policy for general flexible products. Koch et al. (2017) take it a step further by developing a dynamic programming approach for the network revenue management problem with flexible products under customer choice behavior (which naturally leads to dynamic control policies).

Most studies on flexible products assume that the seller defines a set of potential execution modes of a flexible product. In contrast to this, Mang et al. (2012) investigate a flexible product for which customers self-select the level of flexibility. We look into both seller-specified flexible products and customer self-selected ones.

In other industries such as air transportation, flexible products have been used to create inferior services that help with price differentiation because they encourage additional demand from segments that otherwise may not have purchased at the regular prices. For example, an airline might operate several flights on the same day between the same origin and destination; a flexible product could allow the airline to put a customer on any one out of these flights (in exchange for granting the customer a significant discount on the regular flight price). In attended home delivery, the reward for granting the retailer more flexibility will be very small; possibly even zero. However, the cost of the uncertainty to a customer who is anyway at home may be very small as well such that even the message that a flexible slot will be more environmentally friendly may suffice to appeal to a customer. Furthermore, customers are known to be very sensitive to delivery charges as quantified by Yang et al. (2016).

In summary, we can build on a growing body of literature on demand management and vehicle routing for attended home delivery, as well as on flexible products. These two concepts have not yet been combined, and indeed the results from the flexible products literature do not carry over directly because future expected vehicle routing implications need to be taken into account.

3. The Dynamic Pricing Model for Delivery Time Slots

In this section, we first present the dynamic time slot pricing problem for attended home delivery services and then formulate the problem as a dynamic program (involving both standard and flexible time slots) under a model of customer delivery slot choice (namely the nested multinomial logit model).

We consider an e-grocer having a fixed number of homogeneous trucks, each with capacity c in terms of homogeneous transport totes. The e-grocer provides delivery services to customers located in non-overlapping areas $a \in \mathcal{A}$ for one fixed delivery day. The delivery slots can be offered from a set of non-overlapping standard slots $\mathcal{S}$, each of the same duration (say, one hour). Flexible slots can be offered from a set $\mathcal{M}$. Each flexible slot m has a certain set of standard slots $\mathcal{S}_m \subseteq \mathcal{S}$ associated with it that the delivery can be assigned to by the retailer. No restriction is imposed on constructing flexible slots. For brevity, we define $\mathcal{F} = \mathcal{S} \cup \mathcal{M}$ as the set of all time slots managed by the retailer.

The dynamic slot pricing problem is modeled by a discrete dynamic program. The problem has T stages denoted by $t = 1, \ldots, T$ corresponding to the time periods in the booking horizon. The final period T denotes the cut-off time after which no more bookings are accepted. We assume that the time periods chosen are sufficiently small such that the probability of more than one request arrival per period is negligible. Each customer belongs to a customer segment $n \in \mathbf{N}$ which is identifiable (i.e. given a customer profile, we can determine the corresponding segment). Customers in the same segment are assumed to choose delivery time slots using the same choice model and, in particular, have the same price sensitivity to delivery charges. For example, we might consider customers who always book working day slots in the early morning or in the evening as members of one particular segment. We assume that an order from a segment-n customer is (on average) worth r_n in profit before delivery costs, and that each order consumes one unit of truck capacity. Let λ represent the probability of a customer arrival in any given time period (the arrival probabilities are assumed to be independent of time only to simplify notation; notice that we can always reduce time-heterogenous arrival rates to a uniform rate by manipulating the underpinning discrete time grid). Given an arrival, μ_a is the likelihood that the requested delivery address is in area a, and η_{an} is the probability that the request is from customer segment n conditional on there being a request from area a.

The system state in stage t is defined by a matrix of accepted orders $\mathbf{x} \in \mathbb{N}^{|\mathcal{A}| \times |\mathcal{F}|}$, and its component x_{as} indicates the number of orders that have been accepted for delivery in time slot s for area a until time t. At every stage t, given state $\mathbf{x}$, we need to make pricing decisions for all feasible delivery time slots when delivery services can be provided. In line with common business practice, we assume prices are chosen from a finite set of potential price points $\mathcal{D} = \{d_\kappa : \kappa \in \mathcal{K} = \{0, 1, \ldots, K\}\}$, where d_0 denotes the null price that drives demand to zero. We need d_0 to model unavailability of a slot. We also assume that a flexible slot is never higher priced than any feasible standard slot; this assumption is not necessarily required, but it seems reasonable if we want to frame standard slots as a superior service offering which therefore comes at a higher charge.

Given accepted orders $\mathbf{x}$, the set of slots in which we can feasibly schedule a delivery in area a is denoted by $\mathcal{F}_a(\mathbf{x})$ consisting of feasible standard slots $\mathcal{S}_a(\mathbf{x})$ and feasible flexible slots $\mathcal{M}_a(\mathbf{x})$. We assume that a flexible slot is feasible as long as it involves at least one feasible standard slot. In practice, this is not an unrealistic assumption; e.g. Tesco is indeed offering their flexible slots even when some of the one-hour regular slots are marked as unavailable. We make this assumption because we are using a very conservative way of checking feasibility. All feasible slots $s \in \mathcal{F}_a(\mathbf{x})$ will be offered at any stage and state (so we do not consider strategically making certain feasible slots unavailable) because we assume that the retailer wants to maximize the number of available options to improve customer satisfaction.

Let us introduce $\mathbf{g} \in \{0, 1\}^{|\mathcal{A}| \times |\mathcal{F}| \times |\mathcal{K}|}$ where $g_{as\kappa} = 1$ represents assignment of price point d_κ to a feasible slot s for any order received from area a, and $g_{as0} = 1$ indicates the assignment of the null

price d_0 to a slot s in area a (which only happens when s is infeasible due to our assumption that all feasible slots are always to be offered to increase customer satisfaction). The action space at state $\mathbf{x}$ is defined as $\mathbf{G}(\mathbf{x}) := \{\mathbf{g} \mid \mathbf{d}_{am}^T \mathbf{g}_{am} \leq \mathbf{d}_{as}^T \mathbf{g}_{as} \, \forall m \in \mathcal{M}_a(\mathbf{x}), \; s \in \mathcal{S}_a(\mathbf{x}), \; a \in \mathcal{A}; \; \sum_{\kappa \in \mathcal{K} \setminus \{0\}} g_{as\kappa} = 1, \forall s \in \mathcal{F}_a(\mathbf{x}), \; a \in \mathcal{A}; \; g_{as0} = 1 \, \forall \, s \notin \mathcal{F}_a(\mathbf{x}), \; a \in \mathcal{A}\}$. Let $C(\mathbf{x})$ denote the minimum cost of delivering orders $\mathbf{x}$; this minimum cost is the outcome of solving a capacitated vehicle routing problem with multiple time windows. We set $C(\mathbf{x}) = \infty$ when there is no feasible solution for the set of orders $\mathbf{x}$.

The transition probability to a new state in the next time period is defined by the probability of a customer arrival combined with the customer's slot selection probability. Customers are faced with more delivery time uncertainty with flexible slots than standard slots when booking their deliveries. Due to this characteristic of flexible slots, we use the nested multinomial logit (nested MNL) model to express the customers' slot selection probability in this paper where the nests indexed by $(\hat{s}, \; \hat{m})$ are defined in terms of flexible and standard slots, respectively.

Let $U_{s\kappa}^n$ denote the utility of booking slot s at price d_κ for a customer from segment n. We compute this utility as $U_{s\kappa}^n = u_{s\kappa}^n + \epsilon_s$ where $u_{s\kappa}^n$ represents a fixed linear predictor function and ϵ_s is a random variable generated from a Gumbel distribution with zero mean. Not booking any time slot is associated with utility u_0^n for a segment-n customer. Let $\omega_{\hat{s}}$ and $\omega_{\hat{m}}$ denote the dissimilarity parameters of standard slot nest $\hat{s}$ and flexible slot nest $\hat{m}$, respectively. The dissimilarity parameter of no-purchase behavior is set to 1. Quoting from Train (2003) (§4.2), the dissimilarity parameter of a nest is a measure of the degree of independence in unobserved utility among the time slots in the nest. The larger the dissimilarity parameter, the less correlation there is between alternatives in the corresponding nest. If the parameter of a nest is 1, all alternatives in the nest are independent (i.e. no correlation). The reader is referred to Train (2003) for further information on the characteristics of the nested logit model. We assume that customers from the same segment have the same price sensitivity towards standard and flexible slots.

Given the prices identified by $\mathbf{g} \in \mathbf{G}(\mathbf{x})$ in area a, the selection probability of standard delivery slot s by a segment-n customer is computed as

$$p_{ans}(\mathbf{g}) = \frac{\mathbf{v}_{ns}^T \mathbf{g}_{as} \left(\sum\limits_{i \in \mathcal{S}_a(\mathbf{x})} \mathbf{v}_{ni}^T \mathbf{g}_{ai} \right)^{\omega_{\hat{s}} - 1}}{\left(\sum\limits_{i \in \mathcal{S}_a(\mathbf{x})} \mathbf{v}_{ni}^T \mathbf{g}_{ai} \right)^{\omega_{\hat{s}}} + \left(\sum\limits_{i \in \mathcal{M}_a(\mathbf{x})} \mathbf{v}_{ni}^T \mathbf{g}_{ai} \right)^{\omega_{\hat{m}}} + v_{n0}}, \tag{1}$$

where $\mathbf{v}_{ns} = \{v_{ns\kappa} = \exp(u_{ns\kappa}/\omega_{\hat{s}}) \mid \forall \, \kappa \in \mathcal{K}\}$ for standard slot $s \in \mathcal{S}_a(\mathbf{x})$, $\mathbf{v}_{nm} = \{v_{nm\kappa} = \exp(u_{nm\kappa}/\omega_{\hat{m}}) \mid \forall \, \kappa \in \mathcal{K}\}$ for flexible slot $m \in \mathcal{M}_a(\mathbf{x})$ and $\mathbf{g}_{as} = \{g_{as\kappa} \mid \forall \kappa \in \mathcal{K}\}$. Note that $v_{ns\kappa}$ is interpreted as the preference weight of slot s priced at d_κ for a segment-n customer. If a customer from area a books slot s, the state $\mathbf{x}$ transitions to state $\mathbf{x} + \mathbf{1}_{as}$.

We can now introduce the value function $V_t(\mathbf{x})$ at state $\mathbf{x}$ in terms of future value functions $V_{t+1}(\mathbf{x})$

as a maximisation problem for action $\mathbf{g}$;

$$V_t(\mathbf{x}) = \max_{\mathbf{g} \in \mathbf{G}(\mathbf{x})} \sum_{a \in \mathbf{A}, n \in \mathbf{N}} \lambda \mu_a \eta_{an} \left(\sum_{s \in \mathcal{F}_a(\mathbf{x})} p_{ans}(\mathbf{g})(r_n + \mathbf{d}^T \mathbf{g}_{as} + V_{t+1}(\mathbf{x} + \mathbf{1}_{as})) + p_{an0}(\mathbf{g}) V_{t+1}(\mathbf{x}) \right). \quad (2)$$

By substituting $p_{an0}(\mathbf{g}) = 1 - \sum_{s \in \mathcal{F}_a(\mathbf{x})} p_{ans}(\mathbf{g})$ in (2) for a customer's order received from segment n in area a, we can rewrite the value function at state $\mathbf{x}$ as follows:

$$V_t(\mathbf{x}) = \max_{\mathbf{g} \in \mathbf{G}(\mathbf{x})} \sum_{a \in \mathbf{A}, n \in \mathbf{N}} \lambda \mu_a \eta_{an} \left(\sum_{s \in \mathcal{F}_a(\mathbf{x})} p_{ans}(\mathbf{g}) \left[r_n + \mathbf{d}^T \mathbf{g}_{as} - \left(V_{t+1}(\mathbf{x}) - V_{t+1}(\mathbf{x} + \mathbf{1}_{as}) \right) \right] \right) + V_{t+1}(\mathbf{x}). \quad (3)$$

Once the booking horizon is finished (i.e., after the cut-off time T), the delivery of accepted orders ($\mathbf{x}$) takes place. Since the company is concerned with the net profit after delivery cost, the boundary condition at stage $T + 1$ given by

$$V_{T+1}(\mathbf{x}) = -C(\mathbf{x}), \quad (4)$$

where $C(\mathbf{x})$ is (as defined above) the minimal cost of delivery all orders $\mathbf{x}$; recall that the cost is defined to be infinite in case that the underpinning vehicle routing problem is infeasible.

4. Pricing Policy under MNL Choice Model

The dynamic program (3)-(4) is intractable because of its large state space. Moreover, computing $C(\mathbf{x})$ in the model is NP-hard since it involves solving a capacitated vehicle routing problem with time windows (Savelsbergh, 1985). Whilst we cannot solve it directly, it is still useful as it motivates the shape of a pricing policy. If we had at least an approximation of the opportunity cost for an order in time slot s in area a as $\Delta_{as}^t(\mathbf{x}) \approx V_{t+1}(\mathbf{x}) - V_{t+1}(\mathbf{x} + \mathbf{1}_{as})$, we should obtain price $\mathbf{g}$ by solving:

$$\arg \max_{\mathbf{g}_{as} \in \mathbf{G}(\mathbf{x})} \sum_{s \in \mathcal{F}_a(\mathbf{x})} p_{ans}(\mathbf{g}) \left[r_n + \mathbf{d}^T \mathbf{g}_{as} - \Delta_{as}^t(\mathbf{x}) \right]. \quad (5)$$

This problem represents the online decision problem: given a customer arrival from area a, state $\mathbf{x}$ of accepted orders until time t, set of feasible delivery slots $\mathcal{F}_a(\mathbf{x})$ in area a, and opportunity costs $\Delta_{as}^t(\mathbf{x})$ for all feasible slots s in area a, we need to obtain the price points for all feasible delivery slots within a very short time period (within a few hundred milliseconds as advised by an industry representative). Thus, an efficient solution of (5) is crucial and depends to a great extent on the structure underpinning the choice model.

Under the nested MNL, this pricing problem is difficult to be solved; however, general attraction models (including MNL as a special case) have strong structural properties that can be exploited in obtaining tractable optimization routines. Accordingly, we propose to fit an MNL model to the data even though a nested MNL model is a better representative of the actual choice behaviour. As our numerical experiments demonstrate, this can lead to good results even though the simulated customer decisions follow a nested MNL model.

11

Given the historical booking data that is generated under an assumption of customers choosing time slots according to a nested MNL model, we can obtain an estimated MNL choice model which approximates the customer choice behavior under the nested MNL choice model. The reader is referred to Yang et al. (2016) for further information about estimation of MNL choice model parameters from transaction data. Let $\hat{u}_{s\kappa}^n$ denote the the utility of booking time slot s at price d_κ for a segment-n customer in the estimated MNL model. Not booking any time slot has the utility $\hat{u}_0^n$ which is normalized to 1. The selection probability of delivery time slot s by a segment-n customer under prices identified by $\mathbf{g} \in \mathbf{G}(\mathbf{x})$ in area a is computed as follows

$$\hat{p}_{ans}(\mathbf{g}) = \frac{\hat{\mathbf{v}}_{ns}^T \mathbf{g}_{as}}{\displaystyle\sum_{j \in \mathcal{F}_a(\mathbf{x})} \hat{\mathbf{v}}_{nj}^T \mathbf{g}_{aj} + 1}, \tag{6}$$

where $\hat{\mathbf{v}}_{ns} = \{\hat{v}_{ns\kappa} = \exp(\hat{u}_{ns\kappa}) \mid \forall \kappa \in \mathcal{K}\}$. Let us drop index a to reduce a notational clutter.

In the remainder of this section, we re-formulate the online pricing problem under the MNL choice model subject to pricing constraints (namely, flexible slots must not be priced higher than any standard slot) using two steps so as to arrive at an equivalent, compact linear programming formulation. In the first step, we model the price dominance constraints in a tractable fashion (more specifically, we write the constraints in such a way that the associated coefficient matrix is totally unimodular. This allows us to solve the combinatorial problem exactly as a linear program (for instance see, Wolsey and Nemhauser (1999)), and thus offers great advantages in solution speed). In the second step, we linearize the objective.
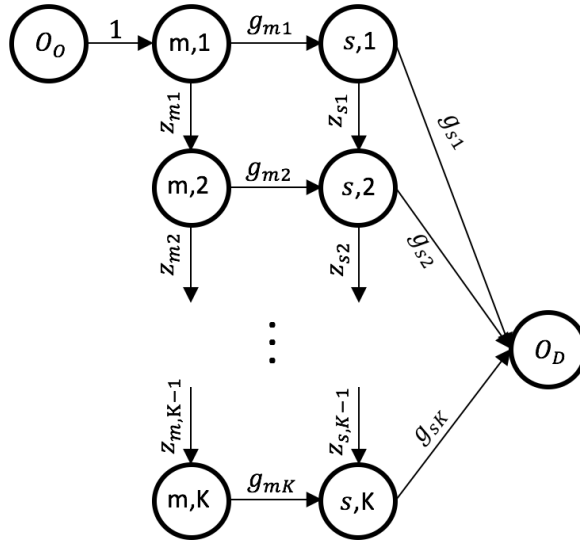

Figure 2: Network structure for flexible slot m and a slot $s \in \mathcal{S}$.

Let us first consider the constraints on admissible prices: prices should be chosen from the discrete set $\mathcal{D}$, and each flexible slot should always be priced no higher than any standard slots since it is an inferior offering (we call the latter price dominance constraints). To formulate the price dominance

constraints in a tractable fashion, we draw on a modeling approach of Davis et al. (2013): they model such price dominance constraints as a unit flow problem on a network because this results in a totally unimodular constraint structure which allows us to relax the integer requirements. To do this, we define a network flow problem for each combination (m, s) of a feasible flexible slot m and one of the feasible standard slots $s \in \mathcal{S}(\mathbf{x})$. There is one source node with unit supply, and one sink with unit demand. Furthermore, we have a node for each combination of m with a price point κ, and likewise for s and each price point κ. The nodes are connected by arcs, as illustrated in Figure 2. Recall that the price points are ordered in increasing value in κ. The flow on some arcs corresponds to pricing variables $g_{m\kappa}$ and $g_{s\kappa}$, and on others we have new variables $z_{j\kappa}$ where $z_{j\kappa} = 1$ if time slot j is priced at d_κ or higher; and 0 otherwise. Enforcing the balance constraints at each node of this network for binary variables g and z ensures that the price for flexible slot m must be less than or equal to the price of slot s. By defining such a network for all (m, s), $m \in \mathcal{M}(\mathbf{x})$, $s \in \mathcal{S}(\mathbf{x})$, we obtain the required constraints to satisfy price dominance with a totally unimodular constraint matrix. The resulting non-linear formulation R_{NLP}^a for a given area a can be stated as follows:

$$R_{\mathrm{NLP}}^a : \max_{\mathbf{g},\mathbf{z}} \quad \sum_{n \in \mathbf{N}} \mu_n \frac{\displaystyle\sum_{s \in \mathcal{F}(\mathbf{x})} \sum_{\kappa \in \mathcal{K}} (r_n - \Delta_s^t + d_\kappa) \hat{v}_{ns\kappa} g_{s\kappa}}{1 + \displaystyle\sum_{s \in \mathcal{F}(\mathbf{x})} \hat{v}_{ns}^T g_s}$$

$$
\begin{aligned}
\text{s.t.} \quad & g_{m1} + z_{m1} = 1, & \forall m \in \mathcal{M}(\mathbf{x}), \\
& g_{m\kappa} + z_{m\kappa} = z_{m,\kappa-1}, & \forall m \in \mathcal{M}(\mathbf{x}), \kappa \in \mathcal{K}\backslash\{0, 1, K\}, \\
& g_{mK} = z_{m,K-1}, & \forall m \in \mathcal{M}(\mathbf{x}), \\
& g_{m1} = g_{s1} + z_{s1}, & \forall m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}), \\
& g_{m\kappa} + z_{s,\kappa-1} = g_{s\kappa} + z_{s\kappa}, & \forall m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}), \kappa \in \mathcal{K}\backslash\{0, 1, K\}. \\
& g_{mK} + z_{s,K-1} = g_{sK}, & \forall m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}), \\
& g_{j\kappa} \in [0, 1], \; z_{j\kappa} \in [0, 1], & \forall j, \kappa \in \mathcal{K}\backslash\{0\}.
\end{aligned}
\tag{7}
$$

The first three groups of constraints state the flow balance at nodes (m, κ) and the next three those of (s, κ). Note that we can drop the binary restrictions since the constraint coefficient matrix of R_{NLP}^a is totally unimodular. This latter property follows from the fact that the balance constraints of a unit network flow problem with single source and sink satisfy the totally unimodular condition Wolsey and Nemhauser (1999).

Proposition 1. *R_{NLP}^a is equivalent to the online pricing problem (5).*

Proof   The proof of this proposition is analogous to the argument provided in Davis et al. (2013). $\square$

Next, we linearize the optimization problem R_{NLP}^a by introducing the following decision variables

for $n \in \mathbf{N}$, $m \in \mathcal{M}(\mathbf{x})$, $s \in \mathcal{F}(\mathbf{x})$ and $\kappa \in \mathcal{K}$

$$\hat{g}_{ns\kappa} = \frac{\hat{v}_{ns\kappa}g_{s\kappa}}{1 + \sum\limits_{j \in \mathcal{F}(\mathbf{x})} \hat{v}_{nj}^T g_j} \quad \text{and} \quad \hat{z}_{nm\kappa} = \frac{z_{m\kappa}}{1 + \sum\limits_{j \in \mathcal{F}(\mathbf{x})} \hat{v}_{nj}^T g_j}.$$

The linear optimization model R_{LP}^a for area a can be formulated as follows:

$$
\begin{aligned}
R_{\mathrm{LP}}^a : \quad & \max_{\hat{\mathbf{g}},\hat{\mathbf{z}}} \quad \sum_{n \in \mathbf{N}} \eta_n \sum_{s \in \mathcal{F}(\mathbf{x})} \sum_{\kappa \in \mathcal{K}} (r_n - \Delta_s^t + d_\kappa)\hat{g}_{ns\kappa} \\
& \text{s.t.} \quad \sum_{s \in \mathcal{F}(\mathbf{x}),\,\kappa \in \mathcal{K}} \hat{g}_{ns\kappa} + \hat{g}_{n0} = 1, && \forall n \in \mathbf{N}, \\
& \qquad \frac{\hat{g}_{nm1}}{\hat{v}_{nm1}} + \hat{z}_{nm1} = \hat{g}_{n0}, && \forall n \in \mathbf{N}, m \in \mathcal{M}(\mathbf{x}), \\
& \qquad \frac{\hat{g}_{nm\kappa}}{\hat{v}_{nm\kappa}} + \hat{z}_{nm\kappa} = \hat{z}_{nm,\kappa-1}, && \forall n \in \mathbf{N}, m \in \mathcal{M}(\mathbf{x}), \kappa = 2, \cdots, K-1, \\
& \qquad \frac{\hat{g}_{nmK}}{\hat{v}_{nmK}} = \hat{z}_{nm,K-1}, && \forall n \in \mathbf{N}, m \in \mathcal{M}(\mathbf{x}), \\
& \qquad \frac{\hat{g}_{nm1}}{\hat{v}_{nm1}} = \frac{\hat{g}_{ns1}}{\hat{v}_{ns1}} + \hat{z}_{ns1}, && \forall n \in \mathbf{N}, m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}) \\
& \qquad \frac{\hat{g}_{nm\kappa}}{\hat{v}_{nm\kappa}} + \hat{z}_{ns,\kappa-1} = \frac{\hat{g}_{ns\kappa}}{\hat{v}_{ns\kappa}} + \hat{z}_{ns\kappa}, && \forall n \in \mathbf{N}, m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}), \kappa = 2, \cdots, K-1, \\
& \qquad \frac{\hat{g}_{nmK}}{\hat{v}_{nmK}} + \hat{z}_{ns,K-1} = \frac{\hat{g}_{nsK}}{\hat{v}_{nsK}}, && \forall n \in \mathbf{N}, m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}), \\
& \qquad 0 \le \hat{\mathbf{g}}, \mathbf{z} \le 1.
\end{aligned}
\tag{8}
$$

The price of slot s that is indicated by the optimal solution $\hat{g}_{nas\kappa}^*$ can be obtained by solving R_{LP}^a. Specifically, it will be priced at d_κ only if $\hat{g}_{nas\kappa}^*$ is non-zero. Note that only one $\hat{g}_{nas\kappa}$ is non-zero among all $\kappa \in \mathcal{K}$ for slot $s \in \mathcal{F}(\mathbf{x})$.

Proposition 2. *Both R_{NLP}^a and R_{LP}^a problems are equivalent and possess the same optimal value.*

Proof The proof is provided in Appendix A. □

Notice that the opportunity cost $\Delta_{as}^t = V_{t+1}(\mathbf{x}) - V_{t+1}(\mathbf{x} + \mathbf{1}_{as})$ for each customer's order coming from area a needs to be estimated for all available slots s at time t. Then, it becomes an input to R_{LP}^a to determine prices of available time slots. As mentioned earlier, it is crucial that the approximation approach must be computationally efficient to cope with the large-scale problems. The next section introduces an approach by adopting a choice-based linear program to estimate related value functions under the estimated MNL choice model.

5. A Model-based Opportunity Cost Approximation

In order to implement the dynamic slotting policy, we require an estimate of the opportunity cost $\Delta_{as}^t(\mathbf{x})$ that is associated with receiving an order from area a at time t in slot s at time t, having currently accepted orders $\mathbf{x}$ on the books. These costs include the marginal cost-to-serve a customer

as well as potential displacement of profit from not being able to satisfy future orders. All this is implicitly captured by the expression $V_{t+1}(\mathbf{x}) - V_{t+1}(\mathbf{x} - \mathbf{1}_{as})$ in the dynamic program (3). Since we cannot calculate the value function $V_t(\mathbf{x})$ exactly, we instead need to approximate it somehow. The challenge here lies in having to account for the vehicle routing problem in the boundary condition (4).

In this section, we will first describe our approximation of the vehicle routing problem. It is designed to yield a cost function that is linear in the number of orders received in a given area and time slot. The linearity of this function is needed so as to ultimately obtain a linear approximation of the value function. In a second step, we can then combine this cost function with a deterministic (non-linear) programming formulation where we assume that future demand is known and equal to its expected value (subject to our chosen future delivery charges). Finally, we exploit the structure of the MNL choice model to equivalently re-formulate this problem into a tractable linear program. The optimal objective of the latter can then be used as an approximation of $V_t(\mathbf{x})$.

Let us start by approximating the final delivery cost as a linear function in the number of orders (already accepted ones as well as expected future ones) per area and time slot. We use the continuous half-width routing method introduced by Daganzo (1987) for this purpose. Since expected future orders in flexible slots need to be assigned to a standard slot, we require the decision variables $\mathbf{w} = \{w_{ams} \mid \forall a \in \mathcal{A},\ m \in \mathcal{M},\ s \in \mathcal{S}_m\}$ that represent the number of accepted orders for flexible slot m that are assigned to standard slot s. Moreover, let $\mathcal{M}_s$ define a set of flexible slots covering standard slot s. The number of orders x'_{as} from area a to be delivered during time slot s is calculated as $x'_{as} = x_{as} + \displaystyle\sum_{m \in \mathcal{M}_s} w_{ams}$.

We assume that only one vehicle is sent to each area; we emphasize that this assumption does not extend to the solution of the full vehicle routing problem at the end of a booking horizon. Instead, this assumption is only used in the opportunity cost approximation where we anyway do not yet have full information on all orders, but where we need to include forecasted orders so as to arrive at reasonable opportunity cost estimates for all slots. Each vehicle has capacity c and delivers customers' orders within a pre-defined rectangular area (with length α_a and width β_a) during the time window of the standard slot. These rectangles can be of different sizes reflecting different densities of customer locations; this approach has been proposed by Yang and Strauss (2017). If the average mile of any vehicle per hour is ν and the average service time per order takes $\bar{\tau}$, then the maximum number of feasible orders B_a to be delivered within an area a for any standard slot s can be calculated as

$$B_a = \frac{t_0 - \frac{2\alpha_a}{\nu}}{\bar{\tau} + \frac{\beta_a}{6\nu}},$$

where t_0 denotes the duration of a standard slot. Given a transportation cost of δ per mile, the delivery cost of orders x'_{as} from area $a \in \mathcal{A}$ and standard slot $s \in \mathcal{S}$ is computed as

$$C_{as}(x'_{as}) = \delta \left(2\alpha_a L(x'_{as}) + \frac{\beta_a}{6} x'_{as} \right),$$

where the function $L(x'_{as})$ is defined as

$$L(x'_{as}) = \begin{cases} 0 & \text{if } x'_{as} = 0, \\ 1 & \text{if } 0 < x'_{as} \leq B_a, \\ \infty & \text{if } x'_{as} > B_a. \end{cases}$$

Note that the cost $\delta\rho_a$ of travelling from the depot to area a (with ρ_a miles distance) is independent from the customer orders, but still contributes to the total delivery cost. We should also mention that these delivery-cost estimations are not used for constructing the final delivery routes.

We are now ready to formulate our non-linear programming approximation of the value function $V_t(\mathbf{x})$. Let $\mathbf{g}' = \{g'_{ias\kappa} \mid \forall i \in \{t, \cdots, T\}, a \in \mathcal{A}, s \in \mathcal{F}, \kappa \in \mathcal{K}\}$ represent pricing decisions made from time t until the end of planning horizon T. The choice probability of a segment-n customer for slot s with price d_κ is computed as

$$p'_{ians\kappa}(\mathbf{g}') = \frac{\hat{v}_{ns\kappa} g'_{ias\kappa}}{\displaystyle\sum_{j \in \mathcal{F}} \hat{\mathbf{v}}_{nj}^T \mathbf{g}'_{iaj} + 1}, \tag{9}$$

where $p'_{ian0}(\mathbf{g}')$ denotes the likelihood of not booking any time slot. Due to notational simplifications, let us define $P_{ians\kappa}(\mathbf{g}') = \lambda\mu_a\eta_{an}p'_{ians\kappa}(\mathbf{g}')$ to denote the probability of a segment-n customer from area a selecting slot s with price d_κ at time i, and accordingly $P_{ian0}(\mathbf{g}') = \lambda\mu_a\eta_{an}p'_{ian0}(\mathbf{g}')$ represents the probability of not booking from a segment-n customer in area a. Then, the expected number of orders to be allocated in standard slot s from area a becomes $\left(x_{as} + \displaystyle\sum_{i=t}^{T} \sum_{n \in \mathbf{N}, \kappa \in \mathcal{K}} P_{ians\kappa}(\mathbf{g}') \right)$.

Given the average profit-before-delivery $(\bar{r}_{an})$ received from the accepted order of segment-n customer from area a, we formulate a non-linear program (V_{NLP}) as follows:

$$(V_{NLP}): \quad \hat{V}_t(\mathbf{x}) = \max_{\mathbf{g}', \mathbf{w}} \sum_{i=t}^{T} \sum_{a \in \mathcal{A}, s \in \mathcal{S}} \sum_{n \in \mathbf{N}, \kappa \in \mathcal{K}} P_{ians\kappa}(\mathbf{g}') (\bar{r}_{an} + d_\kappa) - \sum_{a \in \mathcal{A}, s \in \mathcal{S}} C_{as}(x'_{as})$$

$$\text{s.t.} \sum_{s \in \mathcal{F}} \left[x_{as} + \sum_{i=t}^{T} \sum_{n \in \mathbf{N}, \kappa \in \mathcal{K}} P_{ians\kappa}(\mathbf{g}') \right] \leq c, \qquad \forall a \in \mathcal{A},$$

$$\sum_{s \in \mathcal{S}_m} w_{ams} = x_{am} + \sum_{i=t}^{T} \sum_{n \in \mathbf{N}, \kappa \in \mathcal{K}} P_{ianm\kappa}(\mathbf{g}'), \qquad \forall a \in \mathcal{A}, m \in \mathcal{M}, \tag{10}$$

$$x_{as} + \sum_{i=t}^{T} \sum_{n \in \mathbf{N}, \kappa \in \mathcal{K}} P_{ians\kappa}(\mathbf{g}') + \sum_{m \in \mathcal{M}_s} w_{ams} \leq B_a, \qquad \forall s \in \mathcal{S}, a \in \mathcal{A},$$

$$\sum_{i=t}^{T} \sum_{\kappa \in \mathcal{K}} \frac{P_{ians\kappa}(\mathbf{g}')}{v_{ns\kappa}} = \sum_{i=t}^{T} P_{ian0}(\mathbf{g}'), \qquad \forall a \in \mathcal{A}, s \in \mathcal{F}, n \in \mathbf{N}.$$

$$\mathbf{g}' \in [0, 1], \ \mathbf{w} \geq 0,$$

where x'_{as} is a simplified notation indicating the number of accepted and forecasted expected orders in area a and s. This can be mathematically written as $x'_{as} = x_{as} + \displaystyle\sum_{i=t}^{T} \sum_{n \in \mathbf{N}, \kappa \in \mathcal{K}} P_{ians\kappa}(\mathbf{g}') + \sum_{m \in \mathcal{M}_s} w_{ams}$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}$.

The first set of constraints in (10) ensures that the capacity of vehicles serving in each area is not exceeded. The second group of constraints expresses the balance equations for allocating orders in flexible slots to standard slots while the third set of constraints are time-window constraints for all standard slots after allocating orders from flexible slots. The final set of constraints enforces to have a single price for each time slot requested by a segment-n customer in any area at each time period.

(V_{NLP}) is a difficult optimization problem that involves the nonlinear choice probability terms. We can decompose the problem in terms of areas via our routing cost approximation using independent delivery areas. We build on the ideas of Gallego et al. (2015) to reformulate (V_{NLP}) as a compact linear program. Let $\mathbf{y} = \{y_{asn\kappa} \mid \forall a \in \mathcal{A}, s \in \mathcal{F}, \kappa \in \mathcal{K}\}$ denote new decision variables where $y_{asn\kappa}$ represents the expected number of segment-n customers in area a to select time slot s with price d_κ. Using

$$y_{asn\kappa} = \sum_{i=t}^{T} P_{iasn\kappa}(\mathbf{g}') \text{ and } y_{an0\kappa} = \sum_{i=t}^{T} P_{ian0}(\mathbf{g}'),$$

we obtain the model (V_{LP}) below. Note that the meaning of variables x'_{as} and w_{ams} remains the same. The constraints of the (V_{NLP}) model can be easily transformed into the time-aggregated form as presented in (V_{LP}).

$$(V_{LP}): \quad R_t(\mathbf{x}) = \max_{\mathbf{y},\mathbf{w}} \sum_{a\in\mathcal{A},s\in\mathcal{F}} \sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} (\bar{r}_{an} + d_\kappa)y_{ans\kappa} - \sum_{a\in\mathcal{A},s\in\mathcal{S}} C_{as}(x'_{as}),$$

$$\text{s.t.} \quad \sum_{s\in\mathcal{F}} \left[x_{as} + \sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} y_{ans\kappa} \right] \leq c, \qquad \forall a \in \mathcal{A},$$

$$\sum_{s\in\mathcal{S}_m} w_{ams} = x_{am} + \sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} y_{anm\kappa}, \qquad \forall m \in \mathcal{M}, a \in \mathcal{A},$$

$$x_{as} + \sum_{m\in\mathcal{M}_s} w_{ams} + \sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} y_{ans\kappa} \leq B_a, \qquad \forall s \in \mathcal{S}, a \in \mathcal{A}, \tag{11}$$

$$\sum_{\kappa\in\mathcal{K}} \frac{y_{ans\kappa}}{v_{ns\kappa}} \leq y_{an0}, \qquad \forall a \in \mathcal{A}, s \in \mathcal{F}, n \in \mathbf{N},$$

$$\sum_{a\in\mathcal{A},s\in\mathcal{F}} \sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} y_{ans\kappa} + y_{an0} = \lambda(T-t+1)$$

$$\sum_{\kappa\in\mathcal{K}} d_\kappa \frac{y_{anm\kappa}}{v_{nm\kappa}} \leq \sum_{\kappa\in\mathcal{K}} d_\kappa \frac{y_{ans\kappa}}{v_{ns\kappa}}, \qquad \forall a \in \mathcal{A}, m \in \mathcal{M}, s \in \mathcal{S}, n \in \mathbf{N}.$$

$$\mathbf{y} \geq 0, \ \mathbf{w} \geq 0,$$

where $x'_{as} = x_{as} + \sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} y_{ans\kappa} + \sum_{m\in\mathcal{M}_s} w_{ams}$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}$.

In addition to these constraints, we impose a condition to ensure that total number of customers' bookings over all time slots including no-bookings during the remaining time periods must be equal to the expected number of arrivals. Finally, we have to make sure that any flexible slot is not assigned with a higher price than any standard slots.

Proposition 3. *If the MNL model is considered to describe customer choice behaviour, then (V_{LP}) is equivalent to (V_{NLP}). Thus, $R_t(\mathbf{x}) = \hat{V}_t(\mathbf{x})$.*

Proof   The proof of this proposition is provided in Appendix B.   □

6. Computational Experiments

The central question that we seek to answer with the numerical studies in this section is to what extent, and under what conditions may flexible slots be able to improve profitability? Furthermore, we are interested in quantifying where potential improvements are coming from. Are we saving on routing costs, or attracting more revenue? How are the results affected by varying ratios of demand to capacity? We begin by describing and justifying the scenarios to be analyzed, then report our results and discuss insights and limitations.

6.1. Data and Experimental Design

In our experiments, the delivery day has 14 one-hour non-overlapping standard slots. We focus on a single customer segment and define the utility of booking slot s with price d_κ as $u_{s\kappa} = u_s + \gamma d_\kappa$ where u_s is the utility of the slot and γ indicates the price sensitivity. Table 2 presents those standard slots along with their utility parameters defined under a nested MNL model. We construct 7 flexible slots (labelled as $m_1, m_2, \cdots, m_7$) that can be offered to customers as presented in Table 3 along with their utility parameters under the nested MNL model. Note that the utility parameter of each flexible slot is set as the average utility of its covered standard slots because we assume that customers are not able to anticipate which standard slot they eventually will be assigned to.

Table 2: Utility parameters of standard slots under the nested MNL model

Slot	8–9	9–10	10–11	11–12	12–13	13–14	14–15
u_s	3.2	3.1	3.3	3.2	3.0	2.5	2.7

Slot	15–16	16–17	17–18	18–19	19–20	20–21	21–22
u_s	3.5	3.8	3.9	3.6	4.7	4.2	3.2

$u_0 = 3.5$; $\gamma = -0.45$. Dissimilarity parameters: $\omega_{\hat{s}} = 0.8$ and $\omega_0 = 1$.

We define two scenarios (abbreviated as P3 and A4) for the design of flexible slots in order to test their impact on various performance measures. Scenario P3 features 3 flexible slots (hence '3') covering non-adjacent standard slots with different popularity (hence the 'P'), where popularity refers to the utility of a standard slot excluding delivery charge effects relative to other standard slots in the same flexible slot configuration. The intuition behind this design is that this should allow us to shift some demand from popular slots to less popular ones. Scenario A4 provides four flexible slots (hence the '4') each covering adjacent (hence the 'A') standard slots. This is similar to current industry

practice by Tesco in the UK as they exclusively offer flexible slots consisting of adjacent standard slots. Recall that we are interested in the effect of introducing flexible slots versus not having flexible slots with a nested MNL choice model as the underlying ground truth choice model.

Table 3: Specification of flexible slots and their utility parameters

Time	P3			A4			
Slot	m_1	m_2	m_3	m_4	m_5	m_6	m_7
8–9							
9–10			▓	▓			
10–11			▓	▓			
11–12				▓			
12–13		▓			▓		
13–14	▓				▓		
14–15			▓		▓		
15–16						▓	
16–17		▓				▓	
17–18			▓			▓	
18–19							▓
19–20	▓						▓
20–21		▓					▓
21–22							▓
u_m	3.6	3.7	3.3	3.5	2.7	3.7	4.2

Dissimilarity parameters: $\omega_{\tilde{m}} = 0.5$

As benchmark decision policy, we use dynamic pricing of all 14 standard slots without the ability to offer flexible slots. We report on the performance of being able to use P3 or A4 relative to this benchmark. Note that we also tested a pricing policy based on the nested MNL model. However, it has not significantly improved on the performance measures compared to using the pricing policy with the MNL model (and in some cases even performed worse) and consequently we do not report the corresponding results. It may seem counter-intuitive that using the correct choice model in the policy actually may be worse than using an approximated one; we believe that this effect arises from the fact that our opportunity cost estimation is biased due to the use of MNL in its calculation.

To estimate the MNL-based choice models required by our decision policies in the simulation studies, we generate booking histories involving 320,000 booking requests. We randomly generate for each request a set of standard and flexible slots to represent the historic set of offered alternatives. Each standard time slot has the probability of 70% to be included in the offer set and each flexible slot (either of slots in P3 or A4) is offered when at least one of its covered standards slots is offered. Half of the 320,000 synthetic offer sets were constructed involving flexible slots sampled from P3 and A4, respectively. The price of each offered slot is randomly selected from the set {£4, £5, £6, £7, £8} and flexible slots have prices no higher than standard slots. Specifically, we firstly randomly select

prices for standard slots from the set. Then, we randomly pick prices for flexible slots from a subset consisting of only price points that are lower or equal to the lowest prices of standard slots.

We simulate each customer slot selection decision based on offered slots and their prices following the nested MNL model in Tables 2 and 3 (but this model is not known to our decision policy). Based on our generated booking histories, we use the asclogit package provided in Stata/SE 15 to estimate the parameters of the MNL choice model in Table 4 which are used in the opportunity cost estimation.

Table 4: Utility parameters in the estimated MNL model

Slot	8–9	9–10	10–11	11–12	12–13	13–14	14–15
$\hat{u}_s$	-0.3766	-0.1027	0.2605	-0.2616	-0.73287	-1.3731	-1.1300
Slot	15–16	16–17	17–18	18–19	19–20	20–21	21–22
$\hat{u}_s$	-0.0983	0.2494	0.1509	0.6240	0.7288	0.4977	-0.0984
Flexible	m_1	m_2	m_3	m_4	m_5	m_6	m_7
$\hat{u}_m$	1.0681	0.1282	-1.8204	-0.2369	-2.3910	-0.4260	0.5592

Note: $\hat{u}_0 = 0$; $\hat{\gamma} = -0.5507$.

We focus on a delivery area with the size 15km × 15km with a depot located outside the area at (7km, 16km). The area can be equally divided into 25 sub-areas and customers are evenly distributed within each sub-area. We create a pool of customer locations where 75% of customers is located in the 15 shaded sub-areas and 25% of customers located in the 10 white areas as illustrated in Figure 3. This simple design mimics the situation of a grocery retailer dispatching from a single depot in the outskirts of a city. Customer orders are always of unit size, and their order profit r before delivery cost (excluding delivery charges) is drawn from a normal distribution (truncated at zero) with mean 25 and standard deviation of 10. To justify this choice of mean profit, note that the price of an average grocery basket was about £87 in March 2018 (Retail Gazette (2018)). As stated in Yang et al. (2014), industry practitioners confirmed that a profit margin of 30% is a reasonable rule of thumb for the profit before delivery cost and delivery charges, yielding £26, which we rounded to £25.
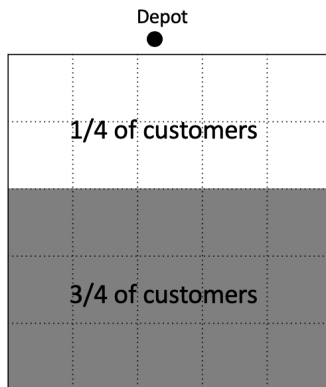

Figure 3: Delivery area

We use the number of time periods covered in one sales horizon to reflect the demand level. We assume that exactly one booking request appears at every time period. Based on the estimated capacity level considering the time window constraints and vehicle capacity, we choose the base demand level with 1800 time periods where the ratio of expected demand to capacity is 1. We apply scaling parameters from the set $\{0.6, \ldots, 1.7\}$ to the demand level to evaluate the performance of our polices under different demand levels. For example, we consider 1980 time periods when the scaling parameter is 1.1.

Each simulation of the sales horizon iterates over all time periods sampling customer slot booking decisions. Slot feasibility checks during the simulation run are performed based on the continuous delivery cost approximation (we stress that this feasibility check is rather conservative). A standard slot is feasible if the vehicle has available capacity and if the number of accepted orders does not exceed B_a. A flexible slot is feasible if and only if at least one of its covered standard slots is deemed feasible. During the booking horizon, opportunity costs are estimated using the approach described in Section 5 and updated after every 100 customer acceptances. When the scaling parameter is 1.0, opportunity costs are re-optimized 12 times during one booking horizon. Slot price points $\mathcal{D}$ range from £2 to £8 in incremental steps of £1.

At the end of the booking horizon, we calculate the delivery costs $C(x)$ by solving a vehicle routing problem with multiple time windows ('multiple' because a flexible slot can be composed by multiple feasible standard slots for a single customer). The company has 25 delivery vans, each with capacity of $c = 100$ units. Travel distance between any two adjacent orders' locations is measured by the Euclidean distance metric. We multiply the total travel distance with a fuel cost of £10 per kilometer to obtain the total delivery costs (we ignore fixed costs). A van travels with a fixed speed of 25km per hour. The service time for each order is 10 minutes.

We apply the simulated annealing approach of Belhaiza et al. (2014) to minimize the total delivery costs. Starting from an initial delivery route, we make iterative improvement steps by randomly reassigning orders within a route. Note that we modify the cost function and the approach proposed by Belhaiza et al. (2014) to evaluate each route by only considering the total delivery costs and penalties from violating the vehicle capacity and time window constraints. Since we use an existing method for constructing routes (and subsequently evaluating costs), we refrain from re-producing the exact algorithm in this paper.

The simulation of each scenario runs 100 times and returns average performance measures on the number of accepted orders, revenue, total delivery costs and total profit. Revenue consists of order revenue and delivery charges. Note that the total profit is computed by subtracting total delivery costs from the total order and delivery charge revenue.

6.2. Numerical Results and Analysis

In our experiments, we aim to measure the value of introducing flexible slots, and to derive insights on what drives this value. The scenarios A4 and P3 only differ in the definition of flexible products. We report the computational results in terms of performance measure determined as percentage change relative to the base scenario of having only standard slots (no flexible products). All other parameters remain the same across all scenarios.

First of all, we are concerned with determining the total profitability impact of P3 versus having no flexible slots in dependence of the demand scaling parameter. Figure 4 demonstrates the percentage increase of profit, which is equal to the product of percentage increase in order volume and percentage increase in mean profit. This product can be approximated by the sum of the same two factors; therefore, we depict the factors as a stacked bar chart so as to give an impression on their relative strength in driving profitability. In all scenarios, adding the three flexible slots increases total profit by at least 3%. Moreover, we can see where these profitability increases are coming from: for low demand scenarios, it is through both attracting more orders and higher profit per order; for high demand scenarios, it is mainly from higher profit per order. Intuitively, what happens under low demand is that more orders are attracted by offering flexible slots to fill up available capacity. Delivery capacity stays constant throughout all experiments, meaning that the larger the scaling parameter, the more congested the delivery routes become. Therefore, we cannot attract many more orders under high demand. However, we may be able to attract more high-value orders.
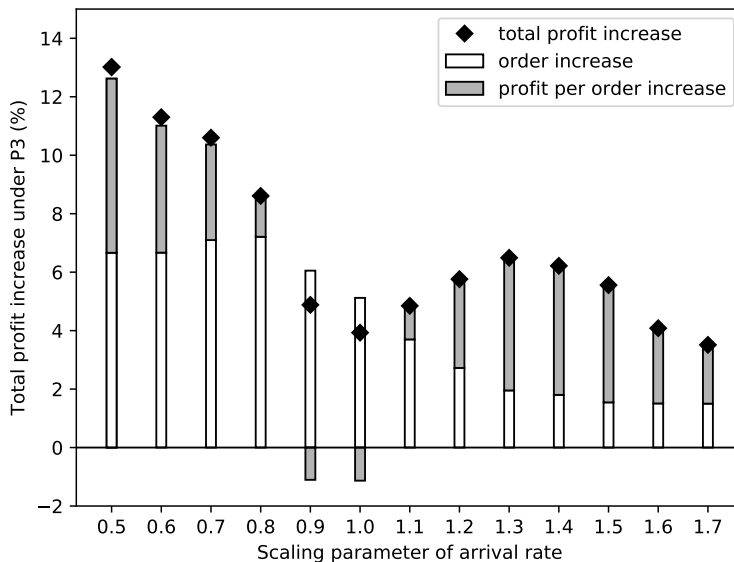

Figure 4: Percentage profit increases under P3 relative to never using flexible slots.

Let us drill down further on the increase of profit per order with the intention of unearthing insights on what causes these profitability improvements. When we consider the profit per order, we observe

that flexible slots significantly increase efficiency regardless of demand levels. We break the profit per order increases further down into percentage changes in revenue per order and percentage changes in cost per order as shown in Figure 5. Let us focus on P3 first; we discuss the comparison with the design A4 further below. Apparently, the main drivers of profit per order increases are routing cost savings at low demand, which demonstrates the value of added flexibility in route planning. Under low demand, routes have relatively few orders to serve, and the ability to move some customers can reduce the length of routes considerably. Accordingly, significant delivery cost savings can be achieved. Based on Figure 6, fleet utilisation improves for low demand scenarios. Under high demand, efficiency of deliveries cannot be improved much because the routes are too congested. We conclude that introducing flexible slots tends to be particularly cost efficient when delivery capacity is large relative to demand.
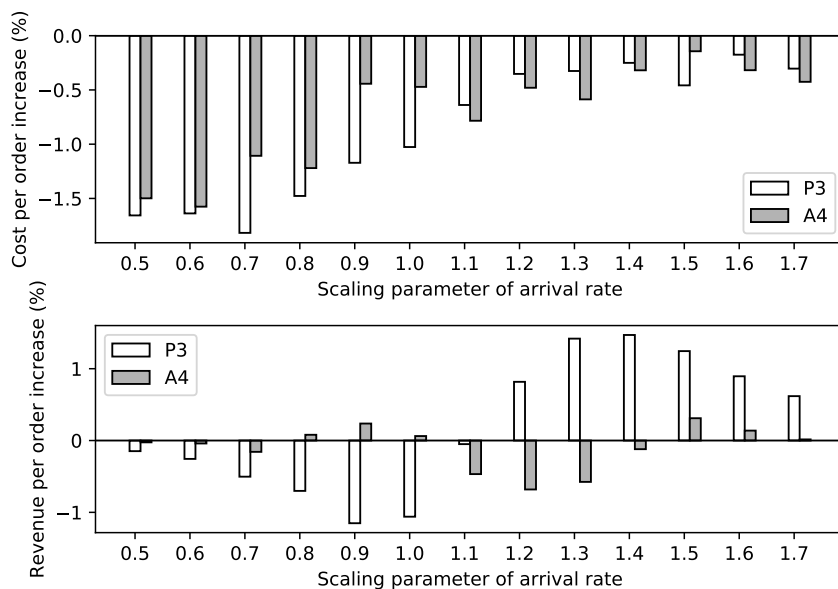

Figure 5: Percentage changes in cost per order and revenue per order under scenarios (A4 versus P3) relative to using only standard slots.
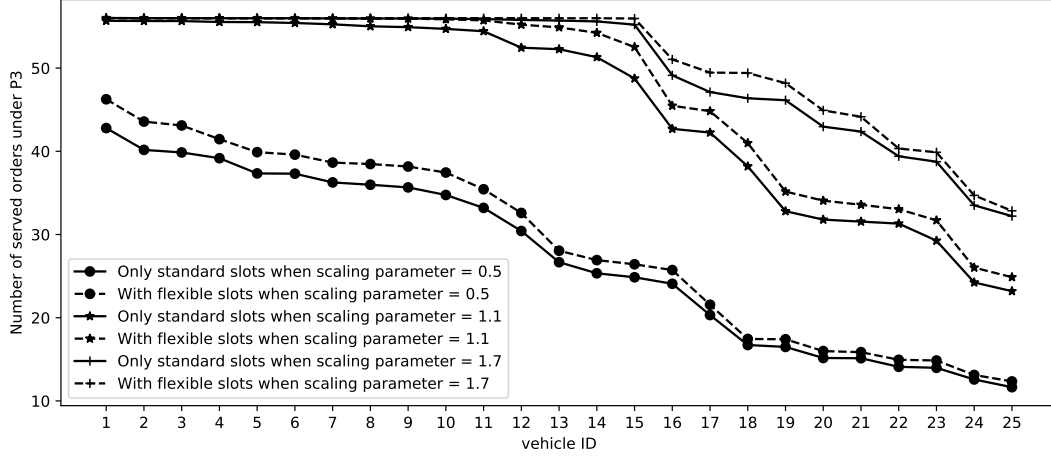
Figure 6: Vehicle utilization under different demand patterns using P3. The utilization is defined as total number of served orders of the vehicle.

On the other hand, the revenue per order decreases in low demand scenarios when we introduce flexible slots. Note that the revenue per order consists of the revenue from the order itself and the delivery service revenue. When demand is low, the available capacity needs to be filled up by attracting more customers with low priced delivery services. Since flexible slots cannot be priced higher than standard slots, the delivery service revenue per order decreases in the low demand scenarios, which results in an overall decrease of revenue per order. When demand is high, the policy focuses on attracting high-value orders and both standard and flexible slots are higher priced resulting in increased revenue per order.

Another interesting question is how to design flexible slots to gain more benefit in reducing delivery costs, i.e., should we group adjacent standard slots together to essentially plan for wider time windows within which the customer is ultimately assigned to a specific standard slot as in scenario A4, or to combine some popular with less popular slots as in scenario P3? We compare these scenarios, A4 versus P3, to obtain some insights to that end. As shown in Figure 7, P3 performs significantly better on profitability in almost all scenarios. When the scaling parameter is small, the increase is mostly driven by additional order intake.
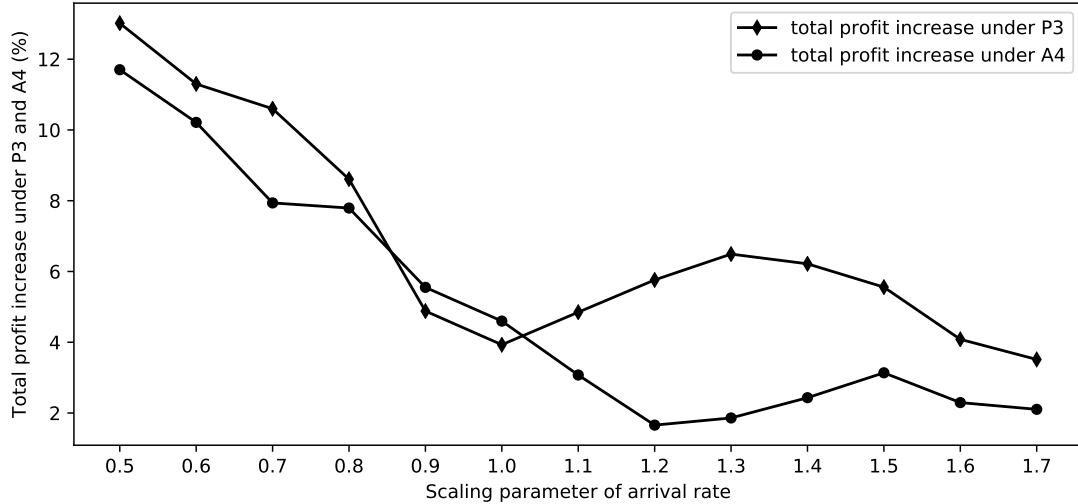
Figure 7: Performance comparison of scenarios (A4 versus P3) relative to using only standard slots.

Both A4 and P3 can reduce cost per order. At low demand levels, P3 results in more cost per order reduction than A4. It makes sense since with P3 we can move customers from popular to less popular slots so as to accommodate more orders on the delivery routes. Therefore, we can conclude that providing wider time slots as flexible slots may not be as efficient as combining popular slots with less popular slots.

Moreover, we are interested in how much customers are paying for delivery services after introducing flexible slots. Figure 8 demonstrates the change in average delivery payment per customer after introducing flexible slots under P3 and A4 scenarios. As demand exceeds the capacity under both scenarios, routes become tight such that it is not benefit of accepting customers in flexible slots. Therefore, our pricing policy charges higher delivery price to customers to control the total demand and prevent them from selecting flexible slots. When demand is lower than the capacity, customers are able to save around 1.5% for the delivery services in P3. This intuitively makes sense as the e-grocer needs to attract customers into unpopular slots by charging low delivery prices. Because flexible slots in P3 are designed as combinations of popular and unpopular standard slots, they are set with low prices to attract customers. However, in the extreme low-demand case (i.e., ratio = 0.5) under P3, e-grocer tends to fit all customers into popular slots by charging unpopular slots with high prices. Because flexible slots in P3 also contain unpopular slots, they are unfavoured by the e-grocer. Accordingly, customers also pay more for delivery services after flexible slots are introduced. On the other hand, since flexible slots in A4 are simply combined with consecutive standard slots, customers can't benefit from the case where e-grocer tries to fill up its unpopular slots using low pricing strategy.
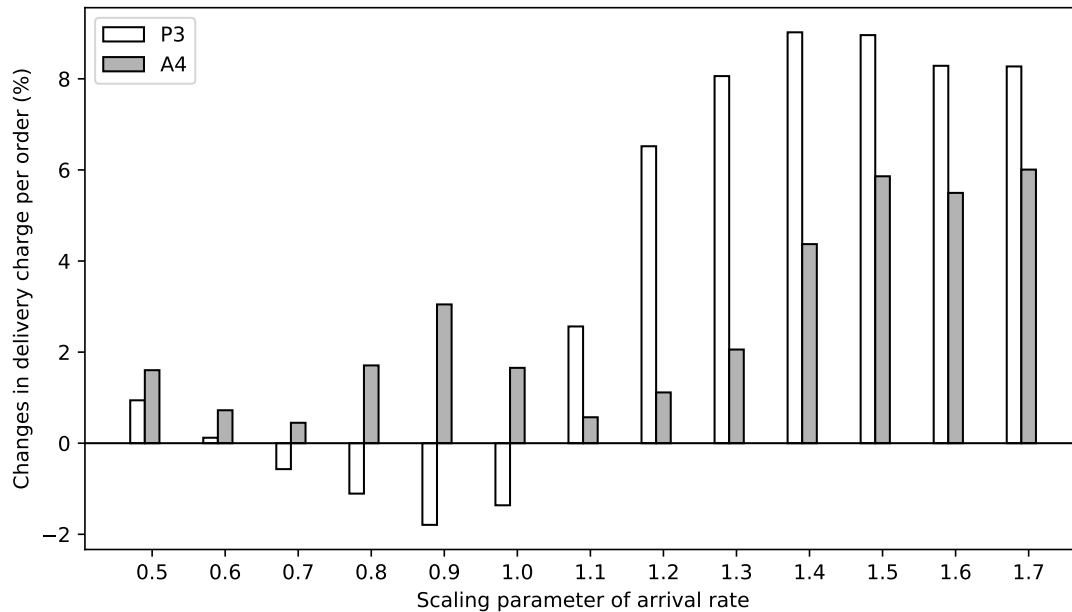
Figure 8: Percentage changes in average delivery charge after offering flexible slots using P3 and A4.

Finally, another interesting practical question is whether customers would typically be assigned to the same mode of a flexible slot (which means that customers' valuations would be affected due to learning effects). To that end, let us classify each standard slot covered by flexible slots as either 'least popular', 'median popular' or 'popular', based on its relative popularity compared to other standard slots covered in the same flexible slots. Note that these labels refer to the utility perception by an average customer and are used as a proxy for the likely capacity tightness. In practice, we would expect that a customer who chose a flexible slot does not view the 'least popular' slot as undesirable; otherwise they would not have chosen this flexible slot. Instead, we expect that such customers would be indifferent to which of the modes they are allocated to — and therefore there should be not dissatisfaction resulting from the allocation decision.

However, one would expect that customers will learn to anticipate their allocation if they typically get assigned the same mode. This would affect their valuation of the flexible slot in as far as its utility would approximate that of the standard slot to which they are always allocated, and it may also lead to gaming behaviour where a customer books the flexible one in the certain expectation of getting the usual assignment mode (which then may lead to disappointment). For that reason, it is desirable to keep the assignment decision to some extent uncertain.

Figure 9 presents final slot allocation decisions for orders in all flexible slots offered in P3. These time slot assignments were made by the routing algorithm at the end of each simulation run; thus, they are not driven by our assumptions on customers' utility parameters of flexible slots. Since popular slots are more congested than other slots, orders in flexible slots are mostly allocated to 'least popular'

26

and 'median popular' slots. As demand is scaled up, the proportion of orders allocated to popular slots increases. When demand is low, our pricing policy tries to retain orders by offering slots at low prices such that a substantial number of customers book into popular standard slots straight away. It results in less delivery (routing) capacity left for accommodating orders in flexible slots. When demand is high, higher prices are charged for slots by our pricing policy, especially for these popular standard slots. It reduces the number of customers who book directly into standard slots but pick to select cheaper flexible slots. Accordingly, more delivery capacity may be found within those popular standard slots such that increases the likelihood of allocating orders in flexible slots to popular standard slots.
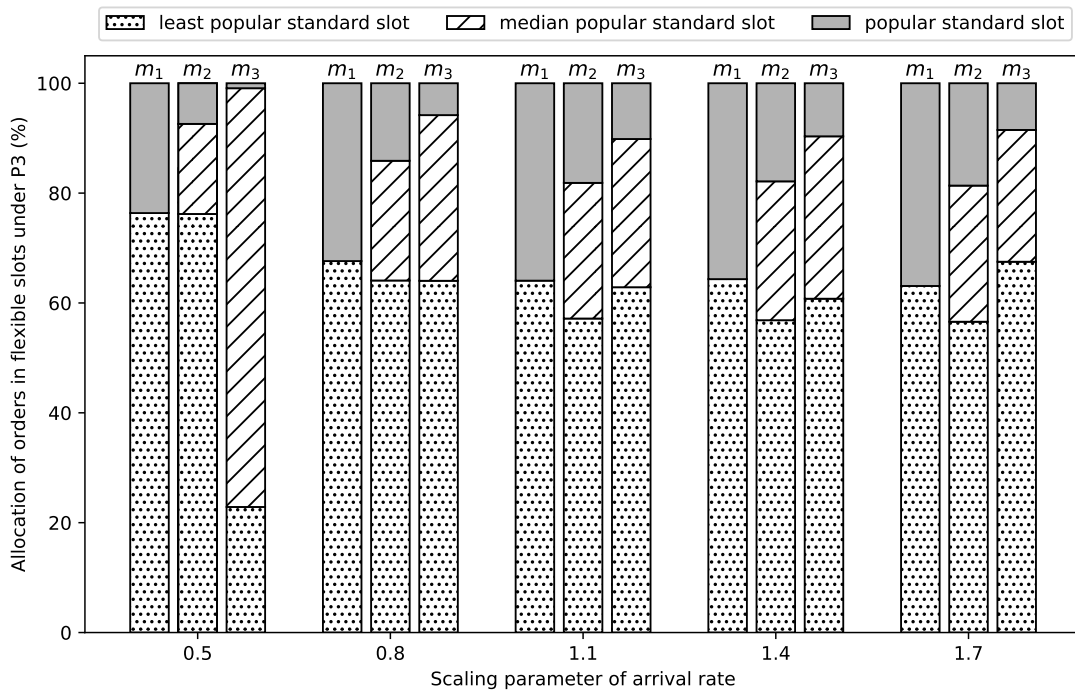
Note: popular slots are 17-18, 19-20, 20-21; median popular slots are 9-10, 16-17; least popular slots are 12-13, 13-14, 14-15

Figure 9: Final allocation for flexible orders under P3

Figure 10 demonstrates final slot allocation decisions for orders in all flexible slots offered under A4. We can observe similar patterns as under P3 that the likelihood of allocating flexible orders to popular standard slots increases as demand level increases, apart from for flexible slot m_5. Flexible slot m_5 includes three less popular standard slots (relative to others on the delivery day). When the demand is low, only a small number of customers books directly into those standard slots and we tend to have delivery capacity left even in popular slots. Therefore, most orders are allocated to the popular standard slot such that the cost per order can be reduced.

Note: popular slots are 10-11, 12-13, 17-18, 19-20; median popular slots are 11-12, 14-15, 16-17, 20-21;
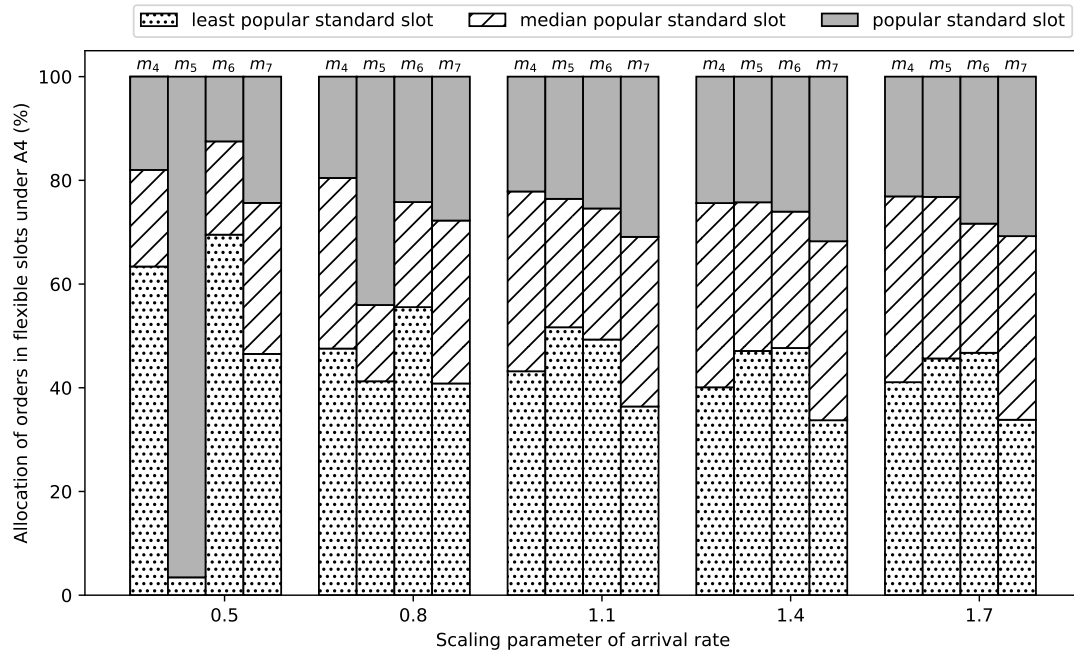least popular slots are 9-10, 13-14, 15-16, 18-19

Figure 10: Final allocation for flexible orders under A4

Moreover, we also observe that orders in flexible slots are more evenly allocated to standard slots under A4 than under P3. Hence, we conclude that a customer would find it harder to anticipate in which slot their order will be executed if flexible slots are simply constructed by combining adjacent standard slots as compared to P3. Still, both designs result in a significant degree of uncertainty for the customer with regard to the final slot allocations, which should ensure that customers will not start to anticipate allocation decisions.

6.2.1. Customizable Flexible Slots

One way of creating a flexible slot is to grant the customer the right to select any two available standard slots (not necessarily adjacent). As for any flexible slot, the retailer would only announce at short notice in which of the two the delivery slots the customer will receive the goods.

To test the impact of introducing a customizable flexible slots requiring exactly two standard slots, we first need to construct a way of modeling demand for such slots. Customers in our simulation experiments choose according to the same nested MNL model as previously, only that we now have all pairs of standard slots in the nest associated with the flexible slots. The utility parameter of each member in this nest is set to the average utility parameter of its two covered standard slots.

In both the online and offline linear programs we use a standard MNL model with as many flexible slots as there are pairs of standard slots. The utility parameters of this MNL model have been estimated from synthetic transaction data using the true nested MNL model; in this manner, we attempt to improve the fit of this simpler choice model. In the online pricing policy, we replace

the constraint that flexible slots are priced no higher than standard slots by the new constraint that all flexible slots have the same price (since they are just possible configurations of the customizable flexible slot priced at a single price). We use the continuous approximation method to check feasibility.

As shown in Table 5, using this 'customizable flexible slot' can provide substantial improvements in overall profit as compared to using standard slots only (all statistically significant at the 95% confidence level).

Table 5: Improvements of using a customizable flexible slot (consisting of exactly two standard slots) alongside standard slots relative to using only standard slots. Profit refers to profit before delivery costs and includes income from delivery charges.

Scale	Number of orders	Total revenue	Total cost	Profit increase (%)	Order increase (%)	Profit/order increase (%)	Cost/order increase (%)
0.5	752	20,538	15,797	28.7	10.7	-0.1	-4.2
0.6	888	24,264	18,101	22	11.1	-0.1	-3.1
0.7	1,008	27,619	20,096	16.8	11.2	-0.4	-2.3
0.8	1,081	30,093	21,140	16.9	9.1	0.4	-2.2
0.9	1,132	31,923	21,867	12.7	7.7	0.1	-1.9
1	1,174	34,045	22,666	11.2	6.7	0.6	-1.1
1.1	1,208	35,795	23,177	12.1	6	1.1	-1.2
1.2	1,209	37,221	23,195	13.6	3.3	3.1	-0.6
1.3	1,217	38,265	23,280	13.9	2.2	4	-0.4
1.4	1,231	39,562	23,539	15.2	1.8	4.8	-0.3
1.5	1,238	40,646	23,471	17.1	1.5	5.5	-0.8
1.6	1,255	41,904	23,789	17.8	1.5	6.1	-0.5
1.7	1,274	43,671	24,161	20.9	1.5	7.6	-0.2

6.2.2. Added Value of Dynamic over Static Slot Pricing (Differently-sized Slots)

So far, we have investigated the added value of flexible slots relative to only using standard delivery time slots (which is standard practice for most UK grocery retailers). Some retailers like Peapod in the US already use flexible slots in the form of offering time windows of different lengths, which can be seen as a special case of the flexible slots considered in this work (namely flexible slots consist of the union of adjacent standard slots). They tend to use a fixed delivery charging scheme, possibly framed in terms of discounts as for Peapod.

Therefore, another interesting question is whether our proposed dynamic pricing concept can add value to such a setting over a static pricing benchmark. As for the latter, we consider delivery charges that depend only on the order value and the length of the time window. Intuitively, we would expect

that the dynamic pricing policy should improve overall profitability since it anticipates the level of future demand (as opposed to the static pricing benchmark) as well as the associated (expected) costs-to-serve.

More specifically, we conduct a set of experiments based on the slot configuration A4 as defined in Table 3. The static pricing policy is defined as follows: for orders with a profit (before delivery costs and excluding delivery charges) of less than or equal to £30, all feasible standard slots are priced at £6. Otherwise, they are priced at £3. All feasible flexible slots are priced £1 below standard slots. All other model parameters are chosen as in our previous experiments under configuration A4.

The results of using this benchmark policy in comparison with our dynamic pricing algorithm are shown in Figure 11 for varying levels of capacity tightness. As expected, we observe that using dynamic pricing consistently leads to improved total profitability across all demand scenarios (all profitability improvements are statistically significant at the 95% level).
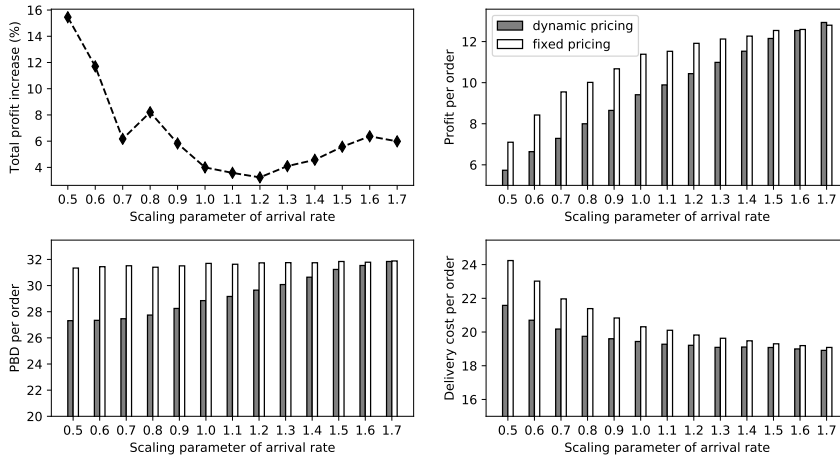

Figure 11: Comparison between dynamic pricing policy and fixed pricing scheme with respect to total profit, profit per order, profit before delivery costs (PBD) per order and delivery costs per order.

The fixed pricing scheme deters more low-value orders; under low demand, this results in higher profit-per-order, but overall poor total profits since we have fewer orders at higher delivery cost per order (a consequence of our inability to steer customers towards cheaper-to-serve slots under static pricing). Under high demand, dynamic pricing focuses on attracting high-value orders due to our ability to anticipate future demand. Still, the ability to influence choice behavior leads to higher profit-per-order and overall improved profitability.

In summary, we conclude that the use of dynamic pricing can be beneficial regardless of demand levels even if a form of flexible slots has already been used. This stems from our ability to influence customers' choice behavior so as to choose cheap-to-serve slots as well as the ability to adjust prioritization of high-value orders in dependence of demand levels.

6.2.3. Illustration: Flexible slots remedy poor feasibility decisions

Our proposed feasibility check is based on a rather conservative continuous routing approximation; it is very quick but tends to be overly cautious. However, using flexible slots can help to remedy the impact of potential poor feasibility decisions as we demonstrate with a small example. We construct this in a way such that it is indeed beneficial to allocate a flexible slot order in a standard slot that was originally deemed infeasible by the continuous routing feasibility check.

In this example, we only have two standard slots (8am-9am and 9am-10am) and one flexible slot (containing both standard slots). Let us assume that using our continuous routing approach to check feasibility tells us that we can accept at most 4 orders in each standard slot. Consider the situation where we have already 4 orders (labelled as A, B, C, D) accepted in the first standard slot, 3 orders (labelled as E, F, G) accepted in the second standard slot. Another customer arrives and chooses the flexible slot (offered because the 9am-10am is deemed feasible, although 8am-9am is not); we label this order as H.

Each order requires 10 minutes service time and the vehicle travels at 25 km/hour. We assume that the delivery costs are measured by the length of the route. The orders are located at the following (X,Y) coordinates: Depot(0.4, -0.5), A(0, 0), B(0.5, 1.1), C(1.8, 0.2), D(2.5, 1.5), E(1, -1), F(3, -0.2), G(2.4, -3), H(0.8, 2). The vehicle departs from the depot at 7am and can reach any location before 8am. Arrival time must be within the promised time window. If order H is allocated in 9am-10am (the one originally deemed feasible), the shortest route is 15.7km as shown in Figure 12. If it is allocated to 8am-9am (originally deemed infeasible), the shortest feasible route is just 12.5km long as shown in Figure 13. This demonstrates that flexible slots can help to remedy poor initial feasibility decisions.
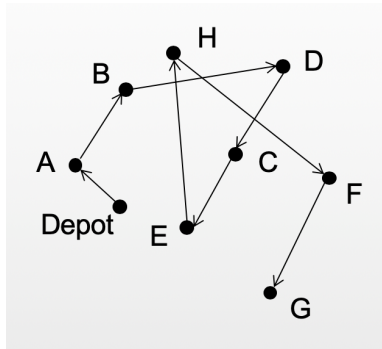

Figure 12: Minimum cost (15.7) route when allocating flexible order H to 9am-10am slot. Order(arrival time, departure time): A(8:00, 8:10), B(8:13, 8:23), D(8:28, 8:38), C(8:42, 8:52), E(9:00, 9:10), H(9:17, 9:27), F(9:34, 9:44), G(9:51, 10:01).
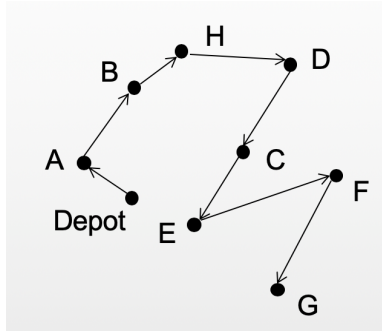
Figure 13: Order(arrival time, departure time): A(8:00, 8:10), B(8:13, 8:23), H(8:26, 8:36), D(8:40, 8:50), C(8:54, 9:04), E(9:08, 9:18), F(9:23, 9:33), G(9:40, 9:50).

6.2.4. Benefits of Single Fully Flexible Slot

A simple way of constructing a flexible slot is to include all standard slots. Since this version of a flexible product would arguably be relatively unattractive for most customers, we assume that its price is fixed at zero. We report in Table 6 the results of offering such a free fully-flexible slot (alongside standard slots)relative to offering only standard slots. The setting otherwise is the same as for our other experiments above.

We observe that even this simple design looks very promising in that it generates substantial profitability increases in almost all scenarios. These are driven by order volume increases and routing cost savings.

Table 6: Simulation results for scenario A1: improvements of offering a single fully-flexible slot (in addition to standard slots) over offering standard slots only. PBD represents profit before delivery costs.

Scaling	Profit increase (%)	Order increase (%)	PBD/order increase (%)	Cost/order increase (%)
0.5	13.1	3.5	0.1	-2.1
0.6	10	4.6	0.1	-1.4
0.7	11.2	4.9	0.2	-1.8
0.8	7.5	5.3	-0.7	-1.8
0.9	3.9	5.2	-1.4	-1.5
1.0	-2	4.9	-3.1	-1.5
1.1	6.4	4.0	0.2	-0.9
1.2	3.2	2.1	0.1	-0.5
1.3	4.1	1.6	0.7	-0.3
1.4	6.6	0.9	1.9	-0.4
1.5	6.5	0.6	2.1	-0.4
1.6	5.1	0.9	1.5	-0.2
1.7	1.8	1.1	0.3	0

6.2.5. Runtime: Feasibility Check, Online and Offline Optimisation

We now illustrate the CPU time taken for the main steps of the proposed procedure: namely feasibility check, the online and offline linear programs. The efficiency is measured in terms of the CPU time taken over a single simulation by using P3 with demand scaling parameter of 1.0. For the run-time of the feasibility check, we compare the efficiency of the insertion heuristics with actual routes to the our continuous approximation procedure.
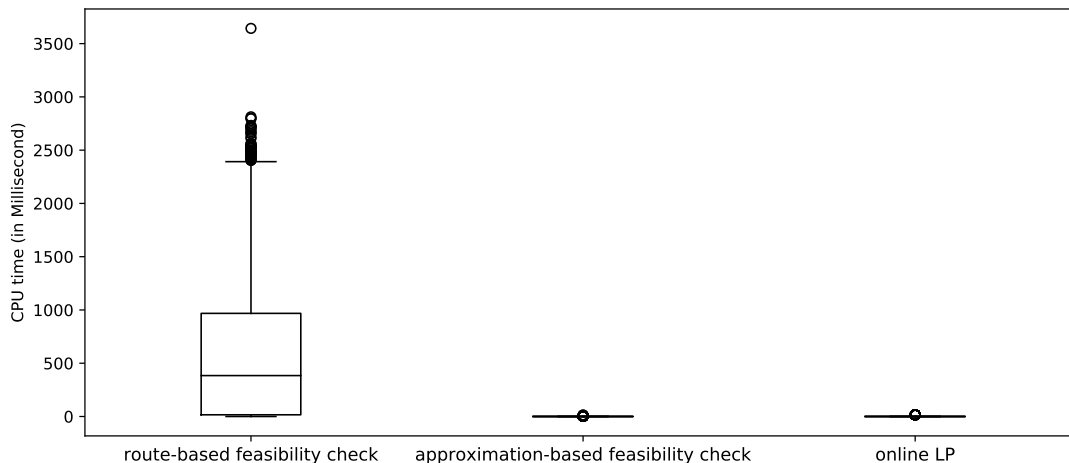

Figure 14: CPU time (millisecond) taken for feasibility check and online linear program

As Figure 14 illustrates, the approximation-based feasibility check and the online LP are essentially solved instantaneous. Checking feasibility based on actual routes is much more time-consuming. In practice, these online calculations should take only a few hundred milliseconds, which means that the route-based feasibility check may not be quick enough for deployment in large-scale online retail environments.

The following table reports the CPU time (millisecond) required to solve the offline linear program which approximates the value function at initial state $(v_0(0))$. We scale a sample problem up in terms

Table 7: CPU time (millisecond) taken to solve the offline problem for various number of segments and areas

Number of	Number of Areas				
Segments	50	100	200	400	800
1	23	57	129	317	787
2	37	103	225	542	1342
4	67	169	385	918	2273
8	115	287	643	1554	3869

of number of segments and areas using the same setting P3 with scale of 1.0. The CPU time increases linearly in the number of segments and in the number of areas, suggesting that the approach is scalable.

The online and offline methods are specifically designed to be tractable. One might wonder why it is important that the offline LP can also be solved very quickly. This is because we want to use it throughout the booking horizon to approximate the opportunity costs for each region and each slot (at a particular point in time). This requires us to solve this LP very frequently – we do not need to do so online (we can just use the most recent available estimates), but should update a look-up table of opportunity cost estimates for each area and slot frequently (ideally after each booking, although in practice this may not be realistic).

7. Conclusions

In this paper, we propose a dynamic pricing approach for standard and flexible time slots for attended home delivery. Flexible slots have recently been introduced by a major retailer in the UK in the form of time windows that encompass four hours; customers who choose such a slot are guaranteed to receive delivery in a one-hour slot within this wider time window. Which slot exactly is communicated only shortly prior to the delivery day. Our method can dynamically price such constructs alongside regular narrow time slots under consideration of the customer choice. The approach is based on tractable linear programming formulations and, as such, is scalable to real-life applications.

Several managerial insights have been obtained via a simulation study. First, flexible slots have significant potential in reducing delivery costs, especially when demand is low relative to available delivery capacity. Moreover, we find that retailers would be better off to construct flexible slots as combinations of some more and some less popular slots, as opposed to the current industry practice of using adjacent slots only. Especially, if demand is high relative to available delivery capacity such flexible slots have the advantage of being able to spread customers more equally across the delivery time slots.

A limitation of the proposed approach is that the solution approach makes use of a rather crude approximation of the capacitated vehicle routing problem with multiple time windows, which forms the boundary condition for the dynamic pricing problem. Nevertheless, as the simulation study shows, flexible products still bring significant routing cost savings (the latter being estimated using an established heuristic taken from the existing literature). More refined approximations may improve results, but at the risk of losing scalability. An interesting future research question is how should these flexible slots be best designed, i.e., which regular slots should be combined to form a flexible slot? Furthermore, how could we price flexible slots when we allow customers to design the flexible slot themselves, i.e., if they can freely combine regular slots to form a custom-made flexible slot?

improve this paper.

Appendix A   Proof of Proposition 2

Consider the nonlinear optimization problem R^a_{NLP} for area a and its optimal solution $(\mathbf{g}^*, \mathbf{z}^*)$. For notational simplicity, we introduce parameter $c_{ns} = r_n - \Delta^t_s$. Let $\mathbf{h} = \{h_{s\kappa} \mid s \in \mathcal{F}(\mathbf{x}), \kappa \in \mathcal{K}\}$ and $\mathbf{f} = \{f_{s\kappa} \mid s \in \mathcal{F}(\mathbf{x}), \kappa \in \mathcal{K}\backslash\{K\}\}$ represent decision variables (corresponding to decisions of the nonlinear optimization model). Given the optimal solution $(\mathbf{g}^*, \mathbf{z}^*)$, we define $V^*_n = \dfrac{\sum_{s,\kappa}(c_{ns} + d_\kappa)\hat{v}_{ns\kappa}g^*_{s\kappa}}{1 + \sum_s \hat{v}^T_{ns}g^*_s}$. The nonlinear problem (7) can then be rewritten as the following linear optimization model:

$$\max_{\mathbf{h},\mathbf{f}} \quad \sum_n \eta_{an} \sum_{s,\kappa}(c_{ns} + d_\kappa - V^*_n)\hat{v}_{ns\kappa}h_{s\kappa}$$

$$\begin{aligned}
\text{s.t.} \quad & h_{m1} + f_{m1} = 1, & \forall m \in \mathcal{M}(\mathbf{x}), \\
& h_{m\kappa} + f_{m\kappa} = f_{m,\kappa-1}, & \forall m \in \mathcal{M}(\mathbf{x}), \kappa \in \mathcal{K}\backslash\{0, 1, K\}, \\
& h_{mK} = f_{m,K-1}, & \forall m \in \mathcal{M}(\mathbf{x}), \\
& h_{m1} = h_{s1} + f_{s1}, & \forall m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}), \\
& h_{m\kappa} + f_{s,\kappa-1} = h_{s\kappa} + f_{s\kappa}, & \forall m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}), \kappa \in \mathcal{K}\backslash\{0, 1, K\}. \\
& h_{mK} + f_{s,K-1} = h_{sK}, & \forall m \in \mathcal{M}(\mathbf{x}), s \in \mathcal{S}(\mathbf{x}) \\
& \mathbf{h}, \mathbf{f} \in [0, 1].
\end{aligned} \tag{12}$$

We first prove that R^a_{NLP} is equivalent to (12), and thus $\sum_n \eta_{an}V^*_n = \sum_n \eta_{an}W^*_n$. The optimal solution of (12) is $(\mathbf{h}^*, \mathbf{f}^*)$ and the objective value is $\sum_n \eta_{an}W^*_n$ where $W^*_n = \sum_{s,\kappa}(c_{ns}+d_\kappa - V^*_n)\hat{v}_{ns\kappa}h^*_{s\kappa}$. Notice that the feasible sets of both problems (7) and (12) consist of the same set of constraints.

- Given the optimal solution $(\mathbf{g}^*, \mathbf{z}^*)$ of R^a_{NLP}, one can easily write the following inequality

$$\sum_n \eta_{an}W^*_n \geq \sum_n \eta_{an} \sum_{s,\kappa}(c_{ns} + d_\kappa - V^*_n)\hat{v}_{ns\kappa}g^*_{s\kappa} \tag{13}$$

  since the optimal solution is also feasible for (12). Furthermore, by substituting

$$V^*_n(1 + \sum_s \hat{v}^T_{ns}g^*_s) = \sum_{s,\kappa}(c_{ns} + d_\kappa)\hat{v}_{ns\kappa}g^*_{s\kappa}$$

  in (13) we obtain

$$\sum_n \eta_{an}W^*_n \geq \sum_n \eta_{an}V^*_n.$$

  This basically implies that the optimal objective value of (12) is at least as large as the optimal objective value of R^a_{NLP}.

- Next, let us consider the optimal solution $(\mathbf{h}^*, \mathbf{f}^*)$ obtained from (12). This is a feasible solution for the problem R^a_{NLP} because both problems have the same search space. Then, we can write

the following valid inequality

$$V_n^* \geq \frac{\sum_{s,\kappa}(c_{ns} + d_\kappa)\hat{v}_{ns\kappa}h_{s\kappa}^*}{\sum_s \hat{v}_{ns}^T h_s^* + 1}$$

that leads to

$$V_n^*\left(\sum_s \hat{v}_{ns}^T h_s^* + 1\right) \geq \sum_{s,\kappa}(c_{ns} + d_\kappa)\hat{v}_{ns\kappa}h_{s\kappa}^* \;\Rightarrow\; V_n^* \geq W_n^*. \tag{14}$$

From this, one can obtain $\sum_n \eta_{an}V_n^* \geq \sum_n \eta_{an}W_n^*$. This shows that the objective value of R_{NLP}^a is at least as large as the optimal objective value of (12).

From these two cases, we find $\sum_n \eta_{an}V_n^* = \sum_n \eta_{an}W_n^*$ that basically states that the optimization problems R_{NLP}^a and (12) are to be equivalent.

Similarly, we can show that R_{LP}^a and (12) are equivalent and they produce the same objective function value. Since V_n^* is a parameter in (12), we can remove it from the objective function. This provides R_{LP}^a by exploiting the structural property of MNL choice model. Let $(\hat{\mathbf{g}}^*, \mathbf{z}^*)$ denote the optimal solution of R_{LP}^a. For $K_n^* = \sum_{s,\kappa}(c_{ns} + d_\kappa)\hat{g}_{ns\kappa}^*$, let us consider the following two cases.

- The optimal solution $(\mathbf{h}^*, \mathbf{f}^*)$ of (12) constructs a feasible solution for R_{LP}^a as

$$\hat{g}_{ns\kappa} = \frac{\hat{v}_{ns\kappa}h_{s\kappa}^*}{(1 + \sum_j \hat{v}_{nj}^T h_j^*)}, \quad \hat{g}_{n0} = \frac{1}{(1 + \sum_j \hat{v}_{nj}^T h_j^*)}, \text{ and } \hat{z}_\kappa = \frac{f_{s\kappa}^*}{(1 + \sum_j \hat{v}_{nj}^T h_j^*)}.$$

Thus, we can then state the following relationship

$$\sum_n \eta_{an}K_n^* \geq \sum_n \eta_{an}\sum_{s,\kappa}(c_{ns} + d_\kappa)\hat{g}_{ns\kappa} = \sum_n \eta_{an}\frac{\sum_{s,\kappa}(c_{ns} + d_\kappa)\hat{v}_{ns\kappa}h_{s\kappa}^*}{1 + \sum_j \hat{v}_{nj}^T h_{aj}^*} = \sum_n \eta_{an}W_n^*. \tag{15}$$

This indicates that the optimal value of R_{LP}^a is greater or equal to the optimal value of (12).

- In the same way, one can show that $(\hat{\mathbf{g}}^*, \hat{\mathbf{z}}^*)$ constructs a feasible solution of the problem (12). In other words, $h_{ns\kappa} = \frac{\hat{g}_{ns\kappa}^*}{(\hat{g}_{n0}^* \hat{v}_{ns\kappa})}$ and $f_{ns\kappa} = \frac{\hat{z}_{ns\kappa}^*}{\hat{g}_{n0}^*}$ is a feasible solution and satisfy the following inequality;

$$\sum_n \eta_{an}W_n^* \geq \sum_n \eta_{an}\sum_{s,\kappa}(c_{ns} + d_\kappa - V_n^*)\hat{v}_{ns\kappa}h_{ns\kappa} = \sum_n \eta_{an}\sum_{s,\kappa}(c_{ns} + d_\kappa - V_n^*)\frac{\hat{g}_{ns\kappa}^*}{\hat{g}_{n0}^*}. \tag{16}$$

Using the relations (15) and $\sum_n \eta_{an}W_n^* = \sum_n \eta_{an}V_n^*$ (as already proven above) in (16), we obtain

$$\sum_n \eta_{an}W_n^* \geq \sum_n \eta_{an}\sum_{s,\kappa}(c_{ns} + d_\kappa)\frac{\hat{g}_{ns\kappa}^*}{\hat{g}_{n0}^*} - \sum_n \eta_{an}\sum_{s,\kappa}K_n^*\frac{\hat{g}_{ns\kappa}^*}{\hat{g}_{n0}^*}$$

$$\geq \sum_n \eta_{an}\frac{K_n^*}{\hat{g}_{n0}^*} - \sum_n \eta_{an}\sum_{s,\kappa}K_n^*\frac{\hat{g}_{ns\kappa}^*}{\hat{g}_{n0}^*} = \sum_n \eta_{an}K_n^*\left(\frac{1}{\hat{g}_{n0}^*} - \sum_{s,\kappa}\frac{\hat{g}_{ns\kappa}^*}{\hat{g}_{n0}^*}\right).$$

Using $\hat{g}_{n0}^* = 1 - \sum_{s,\kappa}\hat{g}_{ns\kappa}^*$, we obtain the following inequality

$$\sum_n \eta_{an}W_n^* \geq \sum_n \eta_{an}K_n^*. \tag{17}$$

This indicates that the optimal value of (12) is not less than the optimal value of R_{LP}^a.

From (15) and (17), we achieve $\sum_n \eta_{an} K_n^* = \sum_n \eta_{an} W_n^*$, and thus R_{LP}^a and (12) are equivalent. In a summary, we can conclude that R_{NLP}^a and R_{LP}^a are equivalent and they possess the same objective value.

Appendix B   Proof of Proposition 3

Suppose that the MNL model is used to describe the customer choice behaviour. We consider the V_{NLP} and V_{LP} problems given state $\mathbf{x}$ at time t. Let $(\mathbf{g}^*, \mathbf{w}_1^*)$ denote the optimal solution of V_{NLP} with the optimal value $\hat{V}^*$. Meanwhile, the optimal solution of V_{LP} is denoted by $(\mathbf{y}^*, \mathbf{w}_2^*)$ and the optimal value is R^*. In order to show that these models are equivalent and produce the same optimal value (i.e., $\hat{V}^* = R^*$) under the MNL choice model, we follow steps in two cases:

Case 1: We first prove that $(\mathbf{g}^*, \mathbf{w}_1^*)$ constructs a feasible solution for V_{LP} so that $R(\mathbf{g}^*, \mathbf{w}_1^*) \leq R^*$. Given $(\mathbf{g}^*, \mathbf{w}_1^*)$, let's define the following decision variables

$$\mathbf{w}' = \mathbf{w}_1^*, \; y'_{asn\kappa} = \sum_{i=t}^{T} P_{ians\kappa}(\mathbf{g}^*) \text{ and } y'_{an0\kappa} = \sum_{i=t}^{T} P_{ian0}(\mathbf{g}^*), \; \forall \kappa \in \mathcal{K}, a \in \mathcal{A}, n \in \mathbf{N}, s \in \mathcal{F}.$$

It can be easily shown that $(\mathbf{y}', \mathbf{w}')$ satisfies the first four sets of constraints in V_{LP}. For the last set of constraints, using $\sum_{a,n} \mu_a \eta_{an} = 1$ and $\sum_{s,\kappa} p'_{ians\kappa}(\mathbf{g}^*) + p'_{ian0}(\mathbf{g}^*) = 1$ in the left-side of the equality, we find

$$\sum_{i=t}^{T} \sum_{a,n} \left(\sum_{s,\kappa} P_{ians\kappa}(\mathbf{g}^*) + P_{ian0}(\mathbf{g}^*) \right) = \lambda \sum_{i=t}^{T} \sum_{a,n} \mu_a \eta_{an} \left(\sum_{s,\kappa} p'_{ians\kappa}(\mathbf{g}^*) + p'_{ian0}(\mathbf{g}^*) \right)$$
$$= \lambda(T - t + 1).$$

This implies that $(\mathbf{g}^*, \mathbf{w}_1^*)$ satisfies all constraints of the V_{LP} model. Therefore, it is a feasible solution for V_{LP} and we can state that $\hat{V}^* \leq R^*$.

Case 2: Let's first reformulate V_{NLP} as a choice-based deterministic linear model (CDLP). Then, using the duality theory, we show that the optimal solution for the dual of $CDLP$ is feasible for the dual problem of V_{LP}. Thus, $R^* \leq \hat{V}^*$ holds.

Let set $\mathcal{G}$ consist of all possible pricing decisions $\mathbf{g}$ for all slots. We define new decision variable $v_{ia}(\mathbf{g})$ to represent the probability of offering price vector $\mathbf{g}$ at time i in area a. For notational convenience, we introduce $P'_{ians\kappa} = \sum_{\mathbf{g} \in \mathcal{G}} P_{ians\kappa}(\mathbf{g}) v_{ia}(\mathbf{g})$ as the decision variable in CDLP which indicates the probability of slot s having price d_κ at time i for a segment-n customer from area a.

Accordingly, CDLP can be formulated as follows:

$$CDLP: \quad \hat{V}_t(\mathbf{x}) = \max_{\mathbf{P}',\mathbf{w}} \sum_{i=t}^{T} \sum_{a\in\mathcal{A},s\in\mathcal{S}} \sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} P'_{ians\kappa}(\bar{r}_{an}+d_\kappa) - \sum_{a\in\mathcal{A},s\in\mathcal{S}} C_{as}(x'_{as})$$

$$\text{s.t.} \sum_{s\in\mathcal{F}}\left[x_{as} + \sum_{i=t}^{T}\sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} P'_{ians\kappa}\right] \le c, \qquad\qquad \forall a\in\mathcal{A},$$

$$\sum_{s\in\mathcal{S}_m} w_{ams} = x_{am} + \sum_{i=t}^{T}\sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} P'_{ianm\kappa}, \qquad\qquad \forall a\in\mathcal{A}, m\in\mathcal{M}, \tag{18}$$

$$x_{as} + \sum_{i=t}^{T}\sum_{n\in\mathbf{N},\kappa\in\mathcal{K}} P'_{ians\kappa} + \sum_{m\in\mathcal{M}_s} w_{ams} \le B_a, \qquad\qquad \forall s\in\mathcal{S}, a\in\mathcal{A},$$

$$\sum_{i=t}^{T}\sum_{\kappa\in\mathcal{K}} \frac{P'_{ians\kappa}}{v_{ns\kappa}} = \sum_{i=t}^{T} P'_{ian0}, \qquad\qquad \forall a\in\mathcal{A}, s\in\mathcal{F}, n\in\mathbf{N}.$$

$$\mathbf{P}' \in [0,1], \ \mathbf{w} \ge 0.$$

Let $\sigma_1 = \{\sigma_{1a} \mid \forall a\in\mathcal{A}\}$, $\sigma_2 = \{\sigma_{2am} \mid \forall a\in\mathcal{A}, \ m\in\mathcal{M}\}$, $\sigma_3 = \{\sigma_{3as} \mid \forall a\in\mathcal{A}, \ s\in\mathcal{S}\}$ and $\sigma_4 = \{\sigma_{4ans} \mid \forall a\in\mathcal{A}, \ n\in\mathbf{N}, \ s\in\mathcal{F}\}$ denote dual decision variables corresponding to constraints of the (primal) $CDLP$. We also introduce $\boldsymbol{\sigma} = \{\sigma_1, \ \sigma_2, \ \sigma_3, \ \sigma_4\}$ for notational simplicity. The constraints of the dual problem of $CDLP$ are

$$\sigma_{1a} + \sigma_{3as} + \frac{\sigma_{4ans}}{\hat{v}_{ans\kappa}} \ge \bar{r}'_{an\kappa}, \qquad \forall a\in\mathcal{A}, s\in\mathcal{S}, n\in\mathbf{N}, \kappa\in\mathcal{K},$$

$$\sigma_{1a} + \sigma_{2am} + \frac{\sigma_{4ams}}{\hat{v}_{anm\kappa}} \ge \bar{r}'_{an\kappa}, \quad \forall a\in\mathcal{A}, m\in\mathcal{M}, n\in\mathbf{N}, \kappa\in\mathcal{K},$$

$$\sigma_{2am} + \sigma_{3as} \ge 0, \qquad\qquad \forall a\in\mathcal{A}, m\in\mathcal{M}, s\in\mathcal{S}_m, \tag{19}$$

$$-\sum_{s\in\mathcal{F}} \sigma_{4ans} \ge 0, \qquad\qquad \forall a\in\mathcal{A}, n\in\mathbf{N},$$

where $\bar{r}'_{an\kappa} = \bar{r}_{an} + d_\kappa - \frac{\delta\beta_a}{6}$ represents the marginal profit-after-delivery in area a with slot price d_κ. The optimal value V_D^* of the dual $CDLP$ problem is achieved at $\boldsymbol{\sigma}^*$.

Similarly, we define dual decision variables $\phi_1 = \{\phi_{1a} \mid \forall a\in\mathcal{A}\}$, $\phi_2 = \{\phi_{2am} \mid \forall a\in\mathcal{A}, \ m\in\mathcal{M}\}$, $\phi_3 = \{\phi_{3as} \mid \forall a\in\mathcal{A}, \ s\in\mathcal{S}\}$, $\phi_4 = \{\phi_{4ans} \mid \forall a\in\mathcal{A}, \ n\in\mathbf{N}, \ s\in\mathcal{F}\}$ and ϕ_5 corresponding to constraints of the primal V_{LP} problem and denote $\boldsymbol{\phi} = \{\phi_1, \ \phi_2, \ \phi_3, \ \phi_4, \ \phi_5\}$. The dual problem of V_{LP} involves the following constraints

$$\phi_{1a} + \phi_{3as} + \frac{\phi_{4ans}}{\hat{v}_{ans\kappa}} + \phi_5 \ge \bar{r}'_{an\kappa}, \qquad \forall a\in\mathcal{A}, s\in\mathcal{S}, n\in\mathbf{N}, \kappa\in\mathcal{K},$$

$$\phi_{1a} + \phi_{2am} + \frac{\phi_{4ams}}{\hat{v}_{anm\kappa}} + \phi_5 \ge \bar{r}'_{an\kappa}, \quad \forall a\in\mathcal{A}, m\in\mathcal{M}, n\in\mathbf{N}, \kappa\in\mathcal{K},$$

$$\phi_{2am} + \phi_{3as} \ge 0, \qquad\qquad \forall a\in\mathcal{A}, m\in\mathcal{M}, s\in\mathcal{S}_m, \tag{20}$$

$$\phi_5 - \sum_{s\in\mathcal{F}} \phi_{4ans} \ge 0, \qquad\qquad \forall a\in\mathcal{A}, n\in\mathbf{N}.$$

R_D^* is obtained by the optimal solution ϕ^* of the dual problem of V_{LP}. Next, from the first two sets of constraints in (19), we define $A^*_{ans\kappa} = \sigma^*_{1a} + \sigma^*_{3as} + \frac{\phi^*_{4ans}}{\hat{v}_{ans\kappa}} - \bar{r}'_{an\kappa}$, and $B^*_{anm\kappa} = \sigma^*_{1a} + \sigma^*_{2am} + \frac{\phi^*_{4ams}}{\hat{v}_{anm\kappa}} - \bar{r}'_{an\kappa}$.

Then, the following relationship holds

$$\sum_{s,\kappa} A^*_{ans\kappa} + \sum_{m,\kappa} B^*_{anm\kappa} \geq 0, \ \forall n \in \mathbf{N}, a \in \mathcal{A}. \tag{21}$$

Notice that ϕ_5 can take any value satisfying (21). Since $\phi_5 \geq 0$ and $\sum_{s \in \mathcal{F}} \sigma^*_{4ans} \leq 0$ (that is obtained from (19)), one can easily observe that $\boldsymbol{\sigma}^*$ and ϕ_5 satisfy constraints in (20). Therefore, $R^*_D \leq V^*_D$ holds such that $R^* \leq V^*$.

References

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

N. Agatz, A. Campbell, M. Fleischmann, and M. Savelsbergh. Time slot management in attended home delivery. Transportation Science, 45(3):435–449, 2011.

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

K. Asdemir, V. S. Jacob, and R. Krishnan. Dynamic pricing of multiple home delivery options. European Journal of Operational Research, 196(1):246–257, 2009.

S. Belhaiza, P. Hansen, and G. Laporte. A hybrid variable neighborhood tabu search heuristic for the vehicle routing problem with multiple time windows. Computers & Operations Research, 52: 269–281, 2014.

D. Bühler, R. Klein, and M. Neugebauer. Model-based delivery cost approximation in attended home services. Computers & Industrial Engineering, 98:78–90, 2016.

A. M. Campbell and M. Savelsbergh. Incentive schemes for attended home delivery services. Transportation science, 40(3):327–341, 2006.

A. M. Campbell and M. W. P. Savelsbergh. Decision support for consumer direct grocery initiatives. Transportation Science, 39(3):313–327, 2005.

C. Cleophas and J. F. Ehmke. When Are Deliveries Profitable? Business & Information Systems Engineering, 6(3):153–163, 2014.

C. F. Daganzo. Modeling distribution problems with time windows. Part I. Transportation Science, 21(3):171–179, 1987.

J. Davis, G. Gallego, and H. Topaloglu. Assortment planning under the multinomial logit model with totally unimodular constraint structures. 2013.

Ecommerce-Foundation. Global B2C e-commerce report 2016. 2016.

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

G. Gallego and R. Phillips. Revenue Management of Flexible Products. Manufacturing & Service Operations Management, 6(4):321–337, 2004. ISSN 1523-4614. doi: 10.1287/msom.1040.0054.

G. Gallego, G. Iyengar, R. Phillips, and A. Dubey. Managing flexible products on a network. 2004.

G. Gallego, R. Ratliff, and S. Shebalov. A general attraction model and sales-based linear program for network revenue management under customer choice. Operations Research, 63(1):212–232, 2015.

J. Gönsch, S. Koch, and C. Steinhardt. Revenue management with flexible products: The value of flexibility and its incorporation into DLP-based approaches. International Journal of Production Economics, 153:280–294, 2014.

A. Hübner, H. Kuhn, and J. Wollenburg. Last mile fulfilment and distribution in omni-channel grocery retailing. International Journal of Retail & Distribution Management, 44(3):228–247, 2016.

R. Klein, M. Neugebauer, D. Ratkovitch, and C. Steinhardt. Differentiated time slot pricing under routing considerations in attended home delivery. Transportation Science, 2017.

S. Koch and R. Klein. Time Window Budgets for Anticipation in Attended Home Delivery. Working Paper, 2017.

S. Koch, J. Gönsch, and C. Steinhardt. Dynamic programming decomposition for choice-based revenue management with flexible products. Transportation Science, 51(4):1046–1062, 2017.

C. Köhler, J. F. Ehmke, and A. M. Campbell. Flexible Time Window Management for Attended Home Deliveries. Omega, 2019.

S. Mang, D. Post, and M. Spann. Pricing of flexible products. Review of Managerial Science, 6(4): 361–374, 2012.

Mintel. Online Grocery Retailing - UK - March 2017. Technical report, 3 2017.

Office for National Statistics. Home workers rates and levels: January to March 2016 and 2017. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/adhocs/008283homeworkersratesandlevelsjanuarytomarch2016and2017, 2017.

A. Petrick, J. Gönsch, C. Steinhardt, and R. Klein. Dynamic control mechanisms for revenue management with flexible products. Computers & Operations Research, 37(11):2027–2039, 2010a.

A. Petrick, C. Steinhardt, J. Gönsch, and R. Klein. Using flexible products to cope with demand uncertainty in revenue management. OR Spectrum, 34(1):215–242, 2010b.

Retail Gazette. Price of average basket of groceries jumps 5%. https://www.retailgazette.co.uk/blog/2018/04/price-average-basket-groceries-jumps-5/, 2018.

M. W. P. Savelsbergh. Local search in routing problems with time windows. Annals of Operations Research, 4(1):285–305, 1985.

K. E. Train. Discrete choice methods with simulation. Number 6-7. Cambridge university press, 2003.

T. Visser and M. Savelsbergh. Strategic Time Slot Management: A Priori Routing for Online Grocery Retailing. 2019.

L. A. Wolsey and G. L. Nemhauser. Integer and combinatorial optimization, volume 55. John Wiley & Sons, 1999.

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

X. Yang, A. K. Strauss, C. S. M. Currie, and R. Eglese. Choice-based demand management and vehicle routing in e-fulfillment. Transportation science, 50(2):473–488, 2014.

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