European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

On the Value of Subscription Models for Online Grocery Retail

Laura Wagner[a,*], Catarina Pinto[b], Pedro Amorim[b]

[a] Catolica Lisbon School of Business and Economics, Palma de Cima, Lisboa 1649-023, Portugal
[b] INESC TEC, Faculdade de Engenharia da Universidade do Porto, R. Dr. Roberto Frias, Porto 4200-465, Portugal

ARTICLE INFO

Article history:
Received 30 April 2019
Accepted 8 May 2020
Available online xxx

Keywords:
E-commerce
online retail
subscription models
difference-in-differences

ABSTRACT

Omnichannel retailers are increasingly introducing subscription-based delivery services. By subscribing to this service and paying fees upfront, customers are entitled to have orders delivered to their home for a given period without paying any extra delivery charge. We analyze the resulting changes in customer behavior from two perspectives:(i) ordering behavior and (ii) delivery preferences. The model is estimated from the online transactional data of a grocery retailer and combines matching and difference-in-differences approaches. We confirm that subscription customers spend more per month and purchase more frequently online than customers without subscriptions. However, this outcome is compromised by shifts towards narrower time slots in the mornings and at night, where slots are requested with less advance notice. When weighing the increased revenue and higher operational costs, we show that subscriptions have a negative impact on a retailer's incremental profit. This remains valid for a wide range of assumptions about (i) the cannibalisation of sales from the retailer's offline business, (ii) picking cost and (iii) delivery cost. To mitigate the impact of subscriptions on retailer profits, we develop a data-driven algorithm that predicts whether certain customers should receive promotions for the subscription plan, rather than it being advertised to all customers. As an extension, we also study whether the addition of a minimum order threshold to subscription plans changes consumer behaviour. We find that this introduction encourages customers to seek more variety and increase their basket size, but does not reduce their order frequency, a phenomena which may be ascribed to cross-selling.

1. Introduction

Subscription-based pricing has been a popular alternative to pay-per-usage charges in many industries. For instance, fitness lovers usually pay a flat-rate monthly membership fee to be granted access to a gym. Similarly, consumers pay streaming services an upfront monthly fee to have access to unlimited movies, series and music, while some innovative online businesses offer subscriptions for the restocking of, for instance, personal hygiene items and shaving products.

More recently, subscriptions became also part of the grocery retail industry and the most common (ancillary) service associated with it is home delivery for online purchases. In this model, upon the payment of a membership fee, customers are eligible for free home delivery during a given period. This raises a number of interesting questions we aim to answer in this research: Does the introduction of a subscription-based delivery service change customers' ordering behavior? If so, to what extent? Is it possible that customers become so reliant on this service that they start dismissing other (omnichannel) options, such as *buy online and pick-up in store*? Could it be that customers, once they subscribe, become more exacting, requiring goods to arrive sooner and at more convenient times? And finally, are customers who subscribe more or less profitable than non-subscribers? The objective of this paper is to document the trade-offs that retailers face when they decide to offer a subscription-based delivery service and to shed light on the extent to which such a commitment can be a double-edged sword for them – a boon for marketing departments, a headache for operations managers.

We joined forces with a large omnichannel grocery retailer, which allows customers to either order online and pick-up-in-store for free, or have their goods delivered to their door for a fee. This home delivery service can be paid for either *per-usage* or *via subscription* – a set-up which enables us to explore how consumers alter their *ordering behavior* and *delivery preferences* once they become subscription customers.

The data we were given bears relation only to online sales, encompassing the 265,349 transactions of 36,792 customers, dating from October 2016 to September 2017. The retailer assigns a unique identifier to each customer, enabling us to access each in-

* Corresponding author.
E-mail addresses: lwagner@ucp.pt (L. Wagner), cppinto@fe.up.pt (C. Pinto), pamorim@fe.up.pt (P. Amorim).

https://doi.org/10.1016/j.ejor.2020.05.011
0377-2217/

dividual's order history. For each transaction, we note (i) whether or not the customer availed themselves of the subscription service, (ii) what products their basket was made up of, (iii) their chosen delivery option (home delivery or pick-up in store) and (iv) (in cases where they chose home delivery) in which time slot they opted to receive their goods at home. The latter probably merits some explanation: it is common for online grocers to offer only a few delivery slots to their customers (Agatz, Campbell, Fleischmann, & Savelsbergh, 2011). However, in our research context, customers are offered hundreds of delivery slots that depend on *many* attributes (slot width, time of day, proximity to delivery, and cost). This particularity enables us to study the changes in delivery preferences of customers who adhered to the subscription-based delivery service.

To disentangle the effect of having a subscription from other possible explanations, we make use of a combined matching and difference-in-differences (DID) approach. To be more precise, we observe whether a customer signs up for the delivery service (treated customers) or not (control customers) and we go on to compare the treated customers' outcomes with those of the controls. The DID approach requires us to observe both treated and control customers prior to and after treatment. In our setting, however, how these periods are defined for control customers is not straightforward, since they never sign up. Our matching procedure overcomes this issue; by successively pairing up customers with similar characteristics until a treated customer subscribes, we not only end up with a clean setting that can be used for the DID procedure, but we also address concerns related to possible self-selection problems. That is, one might expect that treated customers are naturally more attached to the online service before signing up than the control group are. Additionally, any remaining endogeneity concerns– which may arise due to unobservable factors– are taken care of by running a sensitivity analysis similar to Rosenbaum (2010).

Our objective is to analyze the effect of having a subscription on consumers' *online* ordering behavior, most often monitored by marketing and/or sales departments, and delivery preferences, normally the preserve of the operations department.

We find that customers who subscribe end up spending 107€ more *per month* and order 1.56 times more often online: numbers that marketing and sales can be pleased with. However, there is also a decrease in products purchased per order: customers reduce the variety of products in their basket by 6 items, buy a total of 13 fewer products, and spend 23€ less *per order*.

From an operational perspective, however, this overall increase in activity is not without its drawbacks. Subscription customers become more reliant upon the labor- (and cost-) intensive home delivery option rather than opting to pick up their groceries in the store. On top of this, home delivery customers' service requirements are raised: these customers (i) demand their products slightly sooner (ii) choose delivery times more convenient for themselves (which tend to be outside "regular" business hours), (iii) prefer narrower delivery windows and, finally, (iv) select delivery slots which are pricier under the pay-per-usage scheme (making it harder for the company to calculate slot pricing accurately).

As subscriptions have positive repercussions for marketing departments (i.e., increased monthly spend and purchase frequency), but a negative impact on operational performance (i.e., increased operational costs), we aim to discover whether customers who subscribe are more profitable. We find that, on average, they are not.

This being the case, why not restrict access to the subscription to only a limited pool of customers? Based on this idea, we go on to develop an algorithm that is able to select which customer should buy into the plan. In particular, the algorithm learns from only the first purchase incidence of customers in the dataset to predict the incremental gain in profit post-subscription. The out of sample estimates suggest that such a strategy could improve the retailer's profit by an average of 23€ per customer.

As an extension to our paper, we also study how the addition of a minimum purchase threshold affects customer behavior. In particular, we are interested in the company's decision, made during the period of analysis, to gradually roll out a model wherein free delivery was contingent upon a minimum spend (in this case 35€).

The gradual roll-out of this policy, along with the price and duration of the subscription remaining the same, was to our benefit, as it allowed us to isolate the effect of the price threshold on consumer ordering patterns. In line with intuition, we find that consumers using a contingent subscription model, rather than an unlimited one, tend to purchase more products and spend more money whenever they go shopping online. However, purchase frequency and monthly spending are unchanged (insignificant results). Furthermore, our findings suggest that customers subject to a contingent model are more vulnerable to cross-selling than to upselling or pantry loading.

These findings can assist grocery retailers in deciding to what extent subscription plans harm the business performance of their online channel.[1]

The remainder of this paper is organized as follows: In Section 2, we review the literature on shipping pricing models. In Section 3, we define the research hypothesis and conceptual model. Section 4 describes the institutional setup and the data used. Section 5 describes the empirical setting and identification strategy. The impact of signing-up for subscription-based delivery services in ordering behavior, delivery preferences and profit is presented in Section 6. In Section 7, we evaluate the impact of subscriptions on retailers' profit on the online channel. In Section 8, we describe the algorithm developed to determine whether a subscription should be advertised to customers. In Section 9, we compare subscription models with and without a threshold (contingent *vs.* unlimited). The paper concludes in Section 10.

2. Literature review

In this paper, we study how subscribing to delivery services impacts customer behavior, both at the marketing and operational level. Traditionally, ancillary services such as delivery were managed as cost centers, (i.e., with the goal being to minimize cost) (Sainathan, 2018). Shipping directly to clients changed this premise: delivering quickly and reliably increases trust, which has positive repercussions on demand, and is especially important in markets with fierce competition and slim margins (e.g., grocery retail) (Mackert, 2019). To this day, there is still no established "best" delivery policy, with different players in the online retail industry presenting a variety of shipping policies to their clients. While some retailers ship goods at no extra charge (e.g., Zappos), others do so only if customers have ordered a minimum amount (e.g., Amazon delivers orders above $25 for free, while charging $9.99 for orders below that value), and some charge customers every time they want products shipped to their home (e.g., Walmart).

Not surprisingly, authors have been keen to explore how customer purchase behaviour alters as a response to changes in shipping structures. Lantz and Hjort (2013) conduct a randomised ex-

[1] For the sake of caution, the results only apply to the online channel of a retailer, rather than to the retailer as a whole. In particular, our results can be generalized to two situations, one in which a grocery retailer only operates online (such as Instacart, Ocado), and one in which an omni-channel retailer analyses channels in isolation.

periment at an online fashion and beauty retailer and find that a lenient shipping policy can increase a consumers' order frequency while bringing down how much they spend per session. Similar conclusions have been drawn in the online grocery context: reduced shipping fees are associated with increased purchase frequency, but such a policy is likely unprofitable when said reduction is taken into account (Lewis, 2006).

Authors have also explored alternatives to these (linear priced) shipping policies, namely the introduction of a (non-linear) order threshold above which customers receive their products at home without paying shipping charges. Estimating a model in which consumers can choose how to allocate their spending over different product categories, the authors find that the presence of a contingent threshold increases basket sizes while encouraging consumers to meet the minimum order threshold, a result which is echoed by (Cachon, Gallino, & Xu, 2018). While these papers provide evidence that shipping policies indeed alter consumers' order behavior, they do not account for the influence of shipping policies on operational aspects, such as the delivery option and slot selection choices. In contrast to the above-mentioned papers we do so in settings where which shipping policy is applied depends on consumers signing up for a subscription. To the best of our knowledge, the only research investigating the impact of the subscription of delivery services on consumer behavior is Belavina, Girotra, and Kabra (2016). The authors compare the *pay per-order* shipping fee with a subscription-based delivery service using a stylized model and find that subscriptions promote more frequent purchases, smaller basket sizes, and higher overall retailer revenue. While the authors focus on the environmental consequences of such subscription model, they do not analyse the delivery preference choices as we do.

When implementing subscriptions, companies expect to change consumption patterns towards an increase in demand. However, these changes can have unintended operational consequences. Using a queueing model, Cachon and Feldman (2011) show that subscriptions lead to less control over system usage than a per-order pricing scheme, although being more effective as a source of revenue. This increase in congestion is easily observed in gyms, for instance, where high usage levels decrease customer satisfaction (since they are not longer able to take full advantage of the service they paid for). Such facts are corroborated by the research in Gourville and Soman (2002), who also acknowledge that managers commonly oversee the impact on consumption when setting pricing policies. In online grocery retail, although customers cannot visualize the increase in congestion, they can still feel its effects: if the increase in congestion makes the retailer run closer to capacity, the probability of failures increases and, consequently, customer frustration. In contrast to these papers, we focus on operational aspects overlooked in the literature: we empirically measure changes in delivery slot preference and their impact on retailers.

To the best of our knowledge, this is the first paper that connects marketing strategy and operational costs in online subscriptions of free deliveries. Moreover, we show that, on average, customers who subscribe to the free delivery service are less profitable than those that pay per usage. Finally, we build an algorithm which enables the retailer to increase profits by advertising the subscription service to a selected group of customers.

3. Research hypothesis and conceptual model

We study whether subscribing to a delivery service alters consumer interaction with a retailer. In particular, we are interested in understanding the changes in consumer ordering behavior and delivery preferences on the online channel.

3.1. Impact of subscription on ordering behavior

When customers sign up for a subscription plan, they pay a one-time fee in advance that allows them to use a company's delivery services without restrictions for a limited period. That is, no per-usage fee is required. Offering such a pricing scheme for ancillary services may change consumers' demand for the goods they order. We theorize that two factors drive the alterations in consumer ordering behavior when signing up for a delivery service: the subscription fee paid in advance and the non-existing per usage fee during the subscription period.

This subscription fee having been paid might trigger what is called a sunk cost effect: in an effort to recover the money spent, customers use the product or service more often than those who do not have it (Thaler, 1980). We thus expect subscription customers to be less willing to switch channels or shop around once they have signed up. Therefore, such customers should become more loyal to the online channel (channel-specific loyalty as opposed to retailer-specific loyalty) and *spend more* there. This effect was previously described in Lantz and Hjort (2013).

After signing up to the delivery service, consumers no longer have to pay for individual deliveries. Inventory models have, for a long time, captured the fundamental trade-offs between quantity and amount spent when consumers make an order choice. According to the well-established Economic Order Quantity (EOQ) framework (Harris, 1913), consumers tend to purchase less frequently but purchase more goods in one session as the per-usage price increases. By this rationale, subscription customers would purchase *more frequently* and *buy fewer items* per-session. This is in accordance with the existing literature on shipping fees described above (Belavina et al., 2016; Lewis, 2006).

Such an effect is also confirmed in the purchasing behavior under uncertainty literature. When shopping for future consumption, customers need to account for the products they will want to use, i.e., for their future preferences. The longer the time gap between purchase and use, the higher the uncertainty regarding future preferences, and the higher the variety of items purchased. Conversely, if the period between two sequential orders is shorter, customers will order less (Guo, 2010; Simonson, 1990). Thus, customers that buy more often should buy fewer items per purchase, thus also *spending less per transaction*.

Hypothesis 1 (Ordering behavior)**.** Subscription customers

(i) spend more online (per month),
(ii) buy more frequently online,
(iii) buy fewer items per online purchase (decrease assortment size and basket size),
(iv) spend less per online purchase.

3.2. Impact of subscription on delivery preferences

Customers signing up for a delivery subscription service may change not only how they purchase products, but also their expectations regarding the service itself. We focus on two main components of delivery preferences, (i) the choice of delivery option (*i.e.*, home delivery versus picking up in store), and (ii) delivery slot selection, in the case of home delivery customers.

Delivery method. Upon finalizing an order, customers have to choose their preferred delivery method: home delivery or pick-up-in-store. While the latter option is commonly free, the former results in an extra cost for the customer, which may put off customers in selecting this option.

Delivering products directly to customers leads to superior customer service (Campbell & Savelsbergh, 2006), and customers are most likely willing to pay for the extra convenience (Agatz, Campbell, Fleischmann, & Savels, 2008). This service is especially useful for people that cannot go shopping by themselves, due to, for instance, physical disabilities, lack of access to transportation or a busy lifestyle, as well as for people who prefer the convenience of receiving the products they need at home (Agatz et al., 2008; Klein, Neugebauer, Ratkovitch, & Steinhardt, 2017). However, if an alternative free-of-charge option is offered (in this case, pick-up-in-store), customers might prefer not to pay the extra cost. We argue, therefore, that waiving the delivery cost will increase the number of customers choosing the home delivery option.

Hypothesis 2 (Delivery method)**.** Subscription customers are more likely to rely upon the "home delivery" option, as opposed to that of "order online and pick-up-in-store" when ordering online.

Delivery slot selection. Minimizing expected delivery costs is important for online grocery retailers. Offering an efficient last-mile delivery is, thus, fundamental in controlling both their costs and customer satisfaction (Agatz et al., 2011; Yang, Strauss, Currie, & Eglese, 2014). One way to achieve such a goal is to offer a menu of time-slots, so that customers self-select the one that fits their preferences and keeps costs down (convenience- price trade off) (Yang et al., 2014). Our partner offers a wide selection of slots characterized by three features: the delivery slot *width*, *time of day* in which the groceries are expected to arrive, and the time between order and delivery (*proximity*). Naturally, one would expect that the retailer's delivery cost is lower for slots that are (i) wider (i.e., allowing the retailer to deliver at any time during the day), (ii) during regular working hours (avoiding payment of night shifts) and (iii) later (giving the retailer more planning flexibility). This is in contrast to what is most convenient for many customers. For instance, while the retailer is better off when wider slots are selected, narrower slots provide a higher service level to the customer (Agatz et al., 2008). To incentivize consumers to make choices which are less convenient for them, our retail partner offers delivery slots for different prices. Customers who want to reduce the amount they spend on deliveries, may act organize their deliveries well in advance. Such behavior has been documented in the air-travel industry, where customers delay a purchase until the last minute to benefit from possible discount prices (Li, Granados, & Netessine, 2014). Similarly, fashion customers might opt to buy at the end-of-season in order to get clearance prices (Cachon & Swinney, 2011). Indulging in such strategic ordering, though, is no longer worth a customer's time once they have signed up for the subscription plan as it makes a retailer's slot pricing redundant.

Hypothesis 3 (Delivery slot selection)**.** When choosing their online delivery slot, subscription customers

(i) select narrower delivery slots,
(ii) request delivery slots at more "convenient" times of the day (e.g., night times),
(iii) select delivery slots closer to the purchase date,
(iv) select delivery slots that are costlier.

4. Institutional setup and data selection

Institutional setup. To answer our research questions, we partnered with a large omnichannel grocery retailer who not only operate brick-and-mortar stores, but also sell their products online. The focal company is the number one provider of groceries in the country. They have 567 stores and sell roughly 35 thousand SKUs per day via their brick-and-mortar stores, and also sell these same products online. Specifically, the retailer offers all items one would expect to find in a supermarket, such as fresh food, groceries, home appliances and hygiene articles via their online channel–a setting which is fairly standard for multi-channel grocery retailers. Even though the retailer operates multiple channels, in the absence of offline channel data, we exclusively focus on the online side of the business.

Each online order comprises a few steps: first, customers add the desired items to their cart, then they decide whether they want to pick the goods up (free of charge) or want them delivered to their home (at an extra cost). If the latter option is selected, customers have to choose their preferred delivery window from a list of available slots. They can choose from a variety of delivery time slots, which are of varying length and available at different times of day and on different days of the week, with each of these characteristics playing a part in the slots' total price.[2]

Data. The retailer shared with us transaction-level online-channel-only information dated between October 2016 and September 2017. Each customer was assigned a unique key, which allowed us to put together a customer purchase trajectory. The individual-level transaction data provided by our partner included customer-related data (customer ID), order-related data (SKU ID, ID of Subscription plan, quantity of each SKU requested, unit price of SKU, discount applied), and shipping and delivery-related data (delivery slot width, delivery slot time, shipping cost, delivery method, time between order and delivery). From the transaction-level data, we extracted and constructed the following variables:

Customer subscription: At each point in time, we observe whether a customer paid for a subscription.[3]

Customer ordering behavior: Our information on each purchase transaction encompasses the check-out time stamp, the stock-keeping unit (SKU), requested quantity per SKU, unit price and eventual discount applied, which are then used to derive variables that help us better understand customers' *ordering behavior*.

- *Monthly spend (€ /month)*: total amount spent by each customer over a given period of time.
- *Frequency (purchases/month)*: number of times a customer purchased over a given period of time.[4]
- *Assortment size (units/purchase)*: total number of *different* products purchased by each customer per purchase transaction.
- *Basket size (units/purchase)*: total number of products purchased by each customer per purchase transaction.
- *Basket value (€ /purchase)*: total amount spent by each customer per purchase transaction.

Delivery preference: Our data on each customer transaction also includes information about the delivery choices they made. Specifically, we know the delivery method (whether it was delivered to their home address or purchased online and picked up in store), as well as characteristics inherent to the delivery slot of choice. This information is the basis of our *delivery preference* variables.

[2] In order to accommodate all existing (online) demand, our partner decided to oversize the delivery slots, which led to low delivery slot stock-out rates - only 1% of the delivery windows were unavailable at order checkout during the time of our study. Instead of paying the slot price every time a customer chooses the home-delivery option, she can decide to pay a one time subscription fee of 26.90€ which qualifies her for free home-delivery for the next 100 days. Given that the average delivery slot cost is 5.90€, such a subscription can be considered "paid off" after just five deliveries. These subscription plans were advertised to everyone who visited the retailer's website.[3]

[3] In our data set, the subscription plan is identifiable via an additional SKU (with a unique SKU number) only at the time the customer subscribed (as opposed to a flag throughout the entire observation period).

[4] To calculate the outcome variable frequency, we count orders placed from the start of the data set to the treatment time, as well as from treatment time to the end of our data set, which gives us the number of purchases prior to and after treatment. We then divide this variable by the number of months that were used to calculate the total number of purchases. To give an example, suppose a customer subscribed on Jan 26, 2017 but had already placed 17 orders between Oct 1, 2016 and that date. In that case Frequency=17/3.847=4.42 orders/month. The Monthly Spending variable was calculated in a similar way, replacing the number of purchases with the total money spent.

Table 1
Summary statistics

Level	Variable	Mean	Std.Dev.	Min	Max	Observations
Customers	Number of orders (units)	7.21	7.32	2	215	36,792
	Home delivery (units)	6.27	7.21	0	213	36,792
Purchase	Assortment size (units)	31.67	20.51	1	300	265,349
	Basket size (units)	60.34	55.93	0.45	2,373.00	265,349
	Basket value (€)	116.99	100.30	0.00	7,869.35	265,349
Home	**Slot Width**					
Delivery	Small (units)	6.22	6.93	0	186	32,406
Orders	Medium (units)	0.06	0.37	0	25	32,406
	Large (units)	0.83	2.25	0	76	32,406
	Slot Time of Day (units)					
	Morning	0.51	2.30	0	118	32,406
	Lunchtime	0.38	1.56	0	77	32,406
	Afternoon	1.30	3.07	0	72	32,406
	Night	1.36	3.31	0	70	32,406
	Flexible	0.25	1.46	0	76	32,406
	Delivery time (days)	1.85	1.33	0.00	29.00	32,406
	Delivery cost (€)	5.56	1.69	0.00	14.90	32,406

Note: In *home delivery orders* the number of observations consists of all customers who choose the option "home delivery" at least once. The table provides the summary statistics for all customers over the entire observational window.

- *Home delivery (%)*: number of times customer requested home delivery over total number of times a customer purchased goods at the retailer in a given period of time.
- *Slot width (%)*: number of times a particular slot (small/medium/large) is chosen over total number of times a customer requested products be delivered to their home in a given period of time.
- *Time of day (%)*: number of times a particular delivery time period (morning/lunchtime/afternoon/night/flexible) is chosen over total number of times a customer requested products be delivered to their home in a given period of time.
- *Proximity (days)*: the time between order and requested delivery for each purchase transaction.
- *Slot cost (€)*: price of the delivery slot selected by each customer.

Sample selection. Three main requirements guide our data selection procedure:

1. The subscription customers we observed purchased online at least once before and once after the acquisition of their membership, and we remove the observations at the exact time of sign-up.
2. No customer received a delivery for free prior to subscribing.
3. Only the first observed subscription period was taken into account, and we removed all observations taking place after the 100 days was completed.

Criteria 1 ensures that we only include those subscription customers whose pre-subscription and post-subscription time period can be properly identified.

Criteria 2 guarantees that no customer in the retained subset signed up for the delivery service prior to the start of our data set, and also that no one received a one-time free delivery promotion. This helps us ascertain that the only reason for free home delivery was because a customer subscribed during the observation period.[5]

Criteria 3 ensures that each customer is represented only once in our retained subset.[6]

Our final sample comprises 36,792 customers and 265,349 purchase transactions. 32,406 of these customers chose the home delivery option at least once. The summary statistics of the most relevant parameters for all customers over the entire observational window is depicted in Table 1.

5. Empirical method

We aim to understand how signing up to a subscription service shapes both customer ordering behavior and delivery preferences. There are, however, some econometric challenges to overcome. Namely, customers who start a subscription might, prior to doing so, exhibit different behaviors to those who choose not to subscribe. Empirically, we approach this endogeneity concern with a combined matching and difference-in-differences setup similar to Bell, Gallino, and Moreno (0000); Calvo, Cui, and Wagner (0000); Li, Propert, and Rosenbaum (2001).

5.1. Matching

The goal of the matching procedure is to find customers who have very similar observable characteristics *prior* to deciding whether or not to subscribe to the delivery service. We split these customers into two groups: those who subscribe (*treated*), and those who never do (*control*). A successful matching removes any systematic differences between these two groups prior to the decision to subscribe (*the treatment event*). Subsequently, we denote with t the first time a customer subscribes to the delivery service. Each potential candidate and the respective treated customer can then be classified according to the *frequency* of their purchases, average *basket size*, average *basket value*, percentage of *home delivery* option chosen and average delivery *slot costs* prior to time t. We form *pairs* (p) consisting of one treated customer and their control counterpart, who have to have (i) purchased at least once prior to t and at least once after t, (ii) purchased *exactly* as often as the treated customer and (iii) been the *closest* in terms of the remaining characteristics (average *basket size*, average *basket value*, percentage of *home delivery* and average delivery *slot cost*[7]) using

[5] Because the retailer only kept track of the purchase date of subscriptions, we are unable to tell whether a customer purchase incidence with free delivery was due to a) the customer having purchased a subscription prior to the start of the data set but within the eligible 100 days period, or b) her having received a one time promotion. In either case, we should discard these customers.

[6] Some customers seemed to have renewed their subscription plan once it ran out. If we had decided to remove this criterion, we would still have needed to discard all of those who had not purchased under the per-usage scheme at least once prior to subscribing a second or third time.

[7] Neither the average *assortment size* nor *monthly spend* has been used as a matching variable, because they are highly co-linear with basket size and basket value, respectively.

Table 2
Summary statistics - Control and Treatment groups after matching.

Variable	*Treatment group*		*Control group*		Observations	Std. Mean differences (smd)
	Mean	Std.Dev	Mean	Std.Dev		
Monthly spend* (€)	125.891	135.528	125.632	138.522	2,700	0.002
Frequency (times)	1.188	1.123	1.188	1.123	2,700	0
Average assortment size* (units)	33.716	17.020	32.643	17.589	2,700	0.062
Average basket size (units)	56.404	33.609	56.550	33.124	2,700	0.004
Average basket value (€)	111.748	65.735	111.343	65.163	2,700	0.006
Home delivery (percent)	92.028	25.674	91.991	25.795	2,700	0.001
Average slot cost (€)	5.250	1.946	5.520	1.734	2,700	0.146

Note: The table presents the observable variables used in our matching procedure to obtain the closest pairs prior to the treatment event. The delivery slot cost for pick-up-in-store transactions are defined as zero. The standard mean differences are pooled and presented in absolute values. *We additionally display the summary statistic of *monthly spend* and *assortment size*, even though these variables are not explicitly used as a matching variable.

Table 3
Summary statistics - Treatment group prior to and after matching.

Variable	*Overall Treatment group*			*Treatment group*		
	Mean	Std.Dev	Observations	Mean	Std.Dev	Observations
Monthly spend* (€)	166.037	177.926	3,604	125.891	135.528	2,700
Frequency (times)	1.652	1.534	3,604	1.188	1.123	2,700
Average assortment size* (units)	32.652	16.383	3,604	33.716	17.020	2,700
Average basket size (units)	54.977	35.429	3,604	56.404	33.609	2,700
Average basket value (€)	110.144	70.970	3,604	111.748	65.735	2,700
Home delivery (percent)	93.902	22.620	3,604	92.028	25.674	2,700
Average slot cost (€)	4.289	2.448	3,604	5.250	1.946	2,700

Note: The table presents the observable variables used in our matching procedure to obtain the closest pairs prior to the treatment event. The Overall Treatment group consists of all subscription customers prior to matching, while the Treatment group consists of the subsample used in the analysis later. The delivery slot cost for pick-up-in-store transactions are defined as zero. *We additionally display the summary statistic of *monthly spend* and *assortment size*, even though these variables are not explicitly used as a matching variable.

the Mahalanobis distance.[8] This matching procedure has the additional advantage of assigning clear pre- and post treatment periods to each control customer, an issue which is, in the absence of such matching, not straightforward, since these customers never sign up. Ties are broken at random. Also, we ensure that control customers are selected without replacement to allow for sufficient variability.

While the matching procedure identifies the closest pairs, they can still be arbitrarily distant. We thus successively reduce the caliper until the sample is balanced in terms of their average pre-treatment outcome variables.[9] Table 3 shows the descriptive statistics of the matched samples. One can see that the mean differences are less than 0.2 standard deviations away, confirming that the matched pairs are balanced –a requirement established by Cochran (1968) and Imai, King, and Stuart (2008). This procedure results in the selection of 5,400 customers, or 2,700 pairs. In Fig. 4 (cf. Appendix), we additionally plot the distribution of several outcome variables prior to and after treatment for both groups, to reinforce that these groups are similar not only based on their averages but also in terms of their distribution.

Another valid concern is whether the treatment sub-sample used in our further analysis represents the overall treatment group. Table 3 illustrates just this. Assortment size, basket size, basket value and home delivery percentage are fairly similar among these two groups. There is, however, a stark difference in frequency and monthly spend: The overall treatment group purchased more frequently than the subsample used for our study, an observation which is likely rooted in the fact that no control member could be located who purchased the same number of times prior to treatment. A similar argument can be made with respect to monthly spend. We argue that this observation just confirms that the matching procedure successfully removed customers who would have biased our results.

5.2. Difference-in-differences

To study the impact of subscription on consumer behavior, we implemented a difference-in-differences (DID) method. For each customer $i = \{1, \dots, C\}$ in pair $p = \{1, \dots, C/2\}$, we observe a vector of purchase transactions which have either taken place prior to or after treatment time t. We let $TREAT_{p,i} = 1$ if a customer i in pair p belongs to the treated group, and zero otherwise. Similarly, $AFTER_t = 1$ is a dummy variable which indicates whether a purchase transaction has occurred after the treatment time t. Once a subscription is purchased, it is valid for 100 days. We consequently remove any purchase transaction that falls outside of this time period.[10] Our goal is to identify the impact of the explanatory variable (subscription) on consumer behavior outcomes, $\mathrm{OUTCOME}_{p,i,t}$ with the following specification:

$$\mathrm{OUTCOME}_{p,i,t} = \alpha \ \mathrm{TREAT}_{p,i} + \beta \ \mathrm{AFTER}_{p,t} + \gamma \ \mathrm{TREAT}_{p,i} \times \mathrm{AFTER}_{p,t} + \lambda_p + \epsilon_{p,i,t}. \quad (1)$$

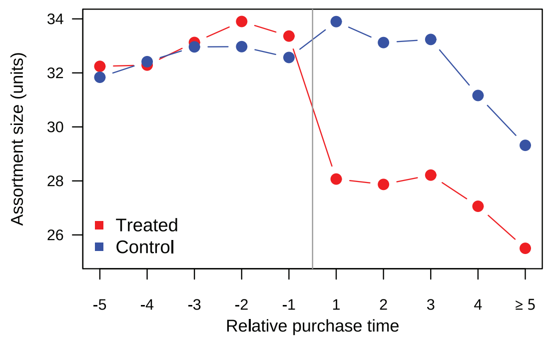
The causal variable of interest, $\mathrm{TREAT}_{p,i} \times \mathrm{AFTER}_{p,t}$, indicates whether a treated customer purchased products while being signed up for a subscription plan, and $\epsilon_{p,i,t}$ is the residual factor. Our regressions include fixed effects at a pair (λ_p) level. Note that we omit time varying fixed effects for our base model because our data set does not have a panel data structure, but is rather defined on a per- purchase transaction level. We do, however, show that all our results are robust when including time fixed effects.

Moreover, we run our estimates using cluster standard errors on a pair level to account for possible correlations among pairs over time. Finally, the dependent variable, $OUTCOME_{p,i,t}$, takes into account a variety of consumer *ordering behavior* (frequency, assortment size, basket size, basket value, monthly spend) and *delivery preference* (home delivery, slot width, time of day, proximity, slot cost), variables previously specified in detail in Section 4.

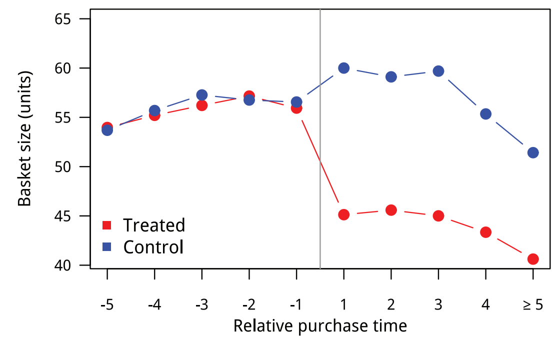
[8] This approach was also followed in Li et al. (2001) and Calvo et al. (0000), which deal with a similar problem.

[9] The specific caliper is 2.83.

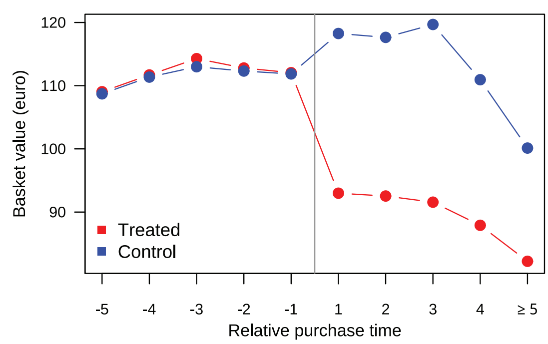
[10] We only look at observations that fall within the minimum of $t + 100$ days and the end of the data period, so effectively the observation time after treatment can end before 100 days.



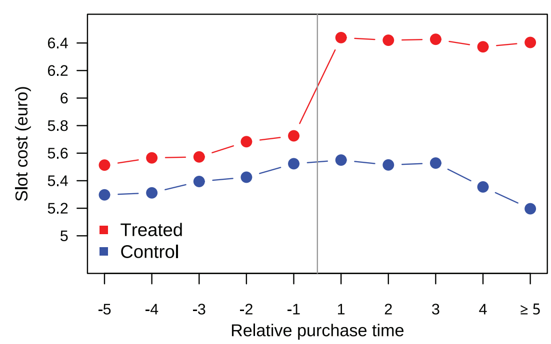
(a) Assortment size over relative purchase time



(b) Basket size over relative purchase time



(c) Basket value over relative purchase time



(d) Slot cost over relative purchase time

Fig. 1. Outcome variables over purchase trajectory

Table 4
The effect of subscription on consumer ordering behavior

	Dependent variable				
	Monthly spend (€ /month)	Frequency (purchases/month)	Assortment size (units/purchase)	Basket size (units/purchase)	Basket value (€ /purchase)
TREAT	0.26	0.00	0.44	- 0.15	0.82
	(0.65)	(0.02)	(0.39)	(0.48)	(0.62)
AFTER	- 14.10***	- 0.21***	0.62**	2.68***	5.61***
	(2.93)	(0.02)	(0.24)	(0.45)	(0.86)
TREAT × AFTER	107.00***	1.56***	- 5.55***	- 12.77***	- 23.44***
	(3.51)	(0.04)	(0.37)	(0.68)	(1.22)
TREAT × AFTER	85.02%	131.6%	−16.86%	- 23.01%	−21.28%
(%)	(78.29,92.31)	(123.4,138.6)	(- 18.25, - 15.36)	(- 24.52, - 21.23)	(- 22.88, - 19.70)
Time-invariant FE	Pair	Pair	Pair	Pair	Pair
Time FE	None	None	None	None	None
Number of customers	5,400	5,400	5,400	5,400	5,400
Observations	10,800	10,800	49,291	49,291	49,291
R^2	0.68	0.65	0.46	0.59	0.59

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. (1) for *monthly spend*, *frequency*, *assortment size*, *basket size* and *basket value*. The first two regressions make use of a two period (prior and after) difference-in-difference models, where the average is calculated as variable/period length (i.e., each customer represents two observations), while the unit of analysis for assortment size, basket size and basket value is a purchase. We cluster standard errors at a pair level for the models *assortment size*, *basket size* and *basket value*. The percentage changes are included with non-parametric bootstrapped 95% confidence intervals (in parentheses) based on 500 random draws. Significance levels: 10% (*), 5% (**), and 1% (***).

Analysis of pre-treatment trends.. Although we matched customers to construct a balanced subsample of pairs, the DID framework additionally requires that treated and control groups exhibit similar trends prior to treatment. Fig. 1 not only illustrates this but also provides a glimpse of what will be rigorously tested later on: After subscription customer sign up their ordering behaviour and delivery preferences change.[11]

[11] We additionally test the parallel trend assumption with an alternative specification similar to Autor (2003) in Section 6.3.

6. Estimation results

In this section, we outline the results of our hypotheses regarding the treatment and moderating effects on the outcome variables.

6.1. Impact of subscriptions on ordering behavior

Main results. Table 4 shows the estimated impact of subscription service membership on consumer ordering behavior, which encompasses the following: frequency, assortment size, basket size,

Table 5
The effect of subscription on consumer delivery option

	Dependent variable home delivery (%)
TREAT	0.04
	(0.39)
AFTER	- 0.60
	(0.39)
TREAT × AFTER	8.04***
	(0.55)
Time-invariant FE	Pair
Time FE	None
Number of customers	5,400
Observations	10,800
R^2	0.72

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. (1) for *home delivery* in percent.

Table 6
The effect of subscription on consumer delivery slot selection

	Hyp. 3(i) Width			*Hyp. 3(ii) Time of day*					*Hyp. 3(iii) Proximity*	*Hyp. 3(iv) Cost*
	Small (%)	Medium (%)	Large (%)	Morning (%)	Lunchtime (%)	Afternoon (%)	Night (%)	Flexible (%)	Proximity (days)	Cost (€)
TREAT	1.08	0.10	- 1.17*	- 1.31*	- 1.63**	- 1.17	6.41***	- 2.30***	- 0.07*	0.20***
	(0.67)	(0.19)	(0.65)	(0.71)	(0.70)	(0.94)	(0.97)	(0.50)	(0.04)	(0.02)
AFTER	- 1.66**	0.92***	0.74	- 0.96	1.27*	- 0.89	0.91	- 0.33	- 0.01	0.08***
	(0.67)	(0.19)	(0.66)	(0.71)	(0.71)	(0.95)	(0.97)	(0.50)	(0.03)	(0.01)
TREAT × AFTER	10.00***	- 0.56**	- 9.43***	6.02***	- 4.91***	- 3.01**	5.67***	- 3.77***	- 0.07*	0.20***
	(0.94)	(0.27)	(0.92)	(1.00)	(0.99)	(1.34)	(1.37)	(0.71)	(0.04)	(0.02)
TREAT × AFTER (%)									- 4.03% (- 7.47, - 0.52)	3.14% (2.61,3.61)
Time-invariant FE	Pair	Pair	Pair	Pair	Pair	Pair	Pair	Pair	Pair	Pair
Time FE	None	None	None	None	None	None	None	None	None	None
Number of customers	5,096	5,096	5,096	5,096	5,096	5,096	5,096	5,096	5,096	5,096
Observations	10,099	10,099	10,099	10,099	10,099	10,099	10,099	10,099	45,284	45,284
R^2	0.42	0.43	0.40	0.35	0.33	0.35	0.42	0.40	0.23	0.41

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Hypothesis 3 - delivery slot selection for delivery slot *length*, *time of day*, *proximity* and *cost*. *Hypothesis 3(i) Length:* the outcome variables are the percentage of *small*, *medium*, and *large* slots selected. Small delivery slot length is defined as less than 3 hours, medium delivery slot length is defined as 3 to 6 hours and large delivery slot length greater 6 hours. *Hypothesis 3(ii) Time of day:* the outcome variables are the percentage of the delivery slots that were selected during the *morning*, *lunchtime*, *afternoon*, and *night* periods, as well as *flexible* slots, *i.e.*, slots that extended for more that one period of time (e.g., morning and lunchtime, or all day). *Hypothesis 3(iii) Proximity:* Standard errors are clustered on a pair basis. *Hypothesis 3(iv) cost:* Purchase online and pick-up-in store observations are removed. Standard errors are clustered at the pair level. The percentage changes for *Hypothesis 3(iii) and 3 (iv)* are included with non-parametric bootstrapped 95% confidence intervals (in parentheses) based on 500 random draws. Significance levels: 5% (*), 1% (**), and 0.1% (***).

basket value and monthly spend. We find that the treatment effect (TREAT × AFTER) is statistically significant for all ordering outcome variables, and that, in general, customers subscribing to the delivery service substantially alter their replenishment patterns. These customers pay out 107€ more per month on average (an increase of 85%), shopping on average 1.56 times more often per month (an increase in activity of 132%). However, they leave with, on average, less 6 distinct products in their assortment (a decrease of 17%), 13 fewer products in their basket (a decrease of 23%), and spend around 23.44€ less per session (a decrease of 21%). As for marketing and sales, the subscription model for delivery services does indeed amplify revenue in the online channel.

6.2. Impact of subscriptions on delivery preferences

Delivery option. Subscription customers are more prone to using the home delivery option. To be precise, they choose home delivery 8.04% more often than those customers who pay per delivery (cf. Table 5). This shift in preferences can be costly for the retailer as it adds more complexity to the already challenging last-mile delivery operation (Agatz et al., 2011).

Delivery slot selection.[12] Having just established that subscription customers use the home delivery service more frequently, we go on to ask how their delivery time slot preferences change. Our goal is, therefore, to understand if customers with a subscription choose more convenient delivery slots (width, time of day and proximity) and if the selected time slots are pricier under the pay-per-usage scheme, as we expected. The answers to these questions are detailed in Table 6.

We find that subscription customers are not as willing to grant the retailer more time to deliver, selecting narrow slots 10% more often than their counterparts - a result which adds unwelcome pressure to the retailer's logistic efficiency (See Column (i) of Table 6).

On top of this, subscription customers prefer having their groceries brought to their address early in the morning or late in the evening (night period), and are less keen to take delivery of their

[12] Given that time slots for home delivery are different from the time slots for pick-up-in-store (in the latter, the slots are wider (solely distinguishing between the morning and afternoon periods) and, therefore, the number of slots available is much lower when compared to the number of slots offered for home delivery), the analysis hereafter focuses only on the home deliveries. Customers that pick-up-in-store were removed, passing from 5,400 customers to 5,096.

Table 7
Outcome variables over customers' purchase trajectory

	Dependent variable			
	Assortment size (units/purchase)	Basket size (units/purchase)	Basket value (€ /purchase)	Slot cost (€ /purchase)
TREAT x $PURCHASE_{-5}$	0.28 (0.96)	- 0.10 (1.72)	- 1.38 (2.84)	0.06 (0.05)
TREAT x $PURCHASE_{-4}$	- 0.25 (0.94)	- 0.87 (1.68)	- 1.37 (2.63)	0.14*** (0.05)
TREAT x $PURCHASE_{-3}$	0.04 (0.92)	- 1.44 (1.60)	- 0.42 (2.41)	0.12*** (0.05)
TREAT x $PURCHASE_{-2}$	0.81 (0.92)	0.02 (1.58)	- 1.23 (2.29)	0.10** (0.05)
TREAT x $PURCHASE_{-1}$	0.67 (0.91)	- 1.00 (1.52)	- 1.51 (2.09)	0.03 (0.04)
TREAT x $PURCHASE_1$	- 5.95*** (0.93)	- 15.25*** (1.62)	- 26.98*** (2.36)	0.25*** (0.05)
TREAT x $PURCHASE_2$	- 5.13*** (0.94)	- 12.63*** (1.62)	- 24.26*** (2.47)	0.28*** (0.05)
TREAT x $PURCHASE_3$	- 4.73*** (0.95)	- 12.75*** (1.69)	- 25.12*** (2.75)	0.24*** (0.05)
TREAT x $PURCHASE_4$	- 4.11*** (0.97)	- 11.01*** (1.79)	- 20.40*** (2.95)	0.28*** (0.05)
TREAT $xPURCHASE_{\geq 5}$	- 4.32*** (0.87)	- 10.77*** (1.61)	- 17.31*** (2.63)	0.26*** (0.05)
Time-invariant FE	Pair	Pair	Pair	Pair
Time-variant FE	None	None	None	None
Number of customers	5,400	5,400	5,400	5,096
Observations	49,291	49,291	49,291	45,284
R^2	0.46	0.59	0.59	0.42

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. (2) for *assortment size*,*basket size*, and *basket value* and *Slot cost* from five purchases prior to treatment to five and more purchases after treatment. Purchase dummies t_{-5} – t_4 are equal to one if a customer ordered t purchases prior or after treatment, while the Purchase dummy $t_{\geq 5}$ is one if that order was the fifth purchase or more after treatment. We suppress the standalone terms. We cluster standard errors at the pair level. Significance levels: 10% (*), 5% (**), and 1% (***).

shopping at lunchtime or during the afternoon. As we can see in Column (ii) of Table 6, subscription customers are 6% more likely to request these times than their pay-per-usage counterparts. These time slots (in particular the evening one), being outside "normal" working hours, are often more labor expensive.

Column (iii) of the same table, meanwhile, shows that subscription customers are also more likely to demand swift delivery. We find the gap between order time and required delivery time is on average 0.07 days, or 4%, shorter, once more requiring the retailer to expend some efforts[13] in service of their customers.

Finally, we conjectured that subscription customers are more likely to select slots for which retailers are charging the most. This hypothesis is indeed confirmed by our findings illustrated in Table 6 Column (iv): treated customers select delivery slots which are, on average, 0.20€ more expensive than those chosen by the control group. Although, at first glance, the magnitude of this difference seems negligible, this finding confirms a consumer tendency to prefer those slots that are operationally more demanding and costly. To conclude, offering subscriptions can invite customers to be far more operationally demanding than they otherwise would be.

6.3. Robustness Checks

In this section, we test whether our results are robust to five possible concerns. Firstly, we look into whether they are sensitive to unobservable biases. Secondly, we rigorously test the parallel trend assumption to complement our graphical illustration in Figure 1. Thirdly, we show that our results remain valid in the face of two alternative specifications of our main model. Finally, we estimate a model which addresses the concern that outcome variables could be correlated.

Sensitivity to unobservable factors. In any quasi-experimental setup where treatment is not perfectly random, treated groups may differ from controls in two ways: They are dissimilar with respect to either observable or unobservable characteristics. Matching treated with control customers addresses the former case, while the manner in which we deal with the latter issue is guided by Rosenbaum (2010). The author recommends performing a sensitivity analysis to find out how strong the unmeasured covariate would have to be in order to change our conclusion. This sensitivity analysis is usually done solely on the post-treatment differences (i.e., neglecting possible differences in pre-treatment periods).

A single parameter, Γ, is employed to capture how much more likely a treated customer is to receive treatment than a control customer. For instance, if $\Gamma = 1$, treated and control customers are equally likely to receive treatment–a setting which represents a perfect randomized experiment. By contrast, if $\Gamma = 2$, treated customers could be twice as likely as controls to subscribe. Thus, low values of Γ indicate that results might be sensitive to hidden biases, while high values suggest the opposite conclusion. In summary, the sensitivity analysis for unobservable factors does not provide evidence that bias is actually present, rather it measures the magnitude of hidden bias that would have to be present before it could alter our conclusions. Table 25 (cf. Appendix) shows the result of this analysis. As we can see, even if our matching procedure had failed to control for unobserved characteristics, the treated group would have to have been more than 22 times (using a p-value of 5%) more likely to buy into the subscription plan before we could conclude that subscription customers do not purchase more frequently. In short, our results with respect to an increase in frequency after subscription are robust to hidden biases. We additionally provide the upper and lower bound of the point

[13] While the results are significant, the magnitude may be practically not as relevant for planning purposes.

Table 8
Comparison of treatment effect between Linear and Multiplicative Models

Model		*Dependent variable* Monthly spend (€ /month)	Frequency (purchases/month)	Assortment size (units/purchase)	Basket size (units/purchase)	Basket value (€ /purchase)
Linear Model	TREAT	0.26	0.00	0.44	- 0.15	0.82
		(0.65)	(0.02)	(0.39)	(0.48)	(0.62)
	AFTER	- 14.10***	- 0.21***	0.62**	2.68***	5.61***
		(2.93)	(0.02)	(0.24)	(0.45)	(0.86)
	TREAT × AFTER	107.00***	1.56***	- 5.55***	- 12.77***	- 23.44***
		(3.51)	(0.04)	(0.37)	(0.68)	(1.22)
	TREAT × AFTER (%)	85.02% (78.29,92.31)	131.6% (123.4,138.6)	−16.86% (- 18.25, - 15.36)	- 23.01% (- 24.52, - 21.23)	−21.28% (- 22.88, - 19.70)
	R^2	0.68	0.65	0.46	0.59	0.59
Log Model	TREAT	0.005	0.00	0.04***	- 0.01	- 0.01
		(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
	AFTER	- 0.02	- 0.06***	0.02**	0.04***	0.05**
		(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
	TREAT × AFTER	0.89***	1.06***	- 0.19***	- 0.25***	- 0.34***
		(0.03)	(0.02)	(0.01)	(0.01)	(0.03)
	TREAT × AFTER (%)	142.29% (130.23,155.51)	189.85% (175.52,205.83)	−17.1% (- 19.05, - 15.27)	- 22.29% (- 23.93, - 20.72)	−29.07% (- 31.75, - 26.40)
	R^2	0.70	0.67	0.38	0.45	0.17
	Time-invariant FE	Pair	Pair	Pair	Pair	Pair
	Time FE	None	None	None	None	None
	Number of customers	5,400	5,400	5,400	5,400	5,400
	Observations	10,800	10,800	49,291	49,291	49,291

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. (1) for *monthly spend*, *frequency*, *assortment size*, *basket size*, and *basket value* in *logarithmic scale*. The first two regressions make use of a two period (before and after) DID model, where the average is calculated as variable/period length (i.e., each customer represents two observations, while the unit of analysis for basket size and basket value is a purchase. The relative change of the Log Model is calculated using Kennedy's approximation (Kennedy, 1981). That is $TREAT \times AFTER = 100 \times (\exp\{x - 1/2V(x)\} - 1)$, where V(x) is the OLS estimate of the variance of c. Cluster standard errors for the last three regressions are taken at a pair level. Significance levels: 10% (*), 5% (**), and 1% (***).

Table 9
Comparision between Model With and Without Monthly Fixed Effects

Model		*Dependent variable* Assortment size (units/purchase)	Basket size (units/purchase)	Basket value (€ /purchase)	Slot cost (€ /purchase)	Proximity (days/purchase)
Without monthly fixed effect	TREAT	0.44	- 0.15	0.82	0.20***	- 0.07*
		(0.39)	(0.48)	(0.62)	(0.02)	(0.04)
	AFTER	0.62**	2.68***	5.61***	0.08***	- 0.01
		(0.24)	(0.45)	(0.86)	(0.01)	(0.03)
	TREAT × AFTER	- 5.55***	- 12.77***	- 23.44***	0.20***	- 0.07*
		(0.37)	(0.68)	(1.22)	(0.02)	(0.04)
	TREAT × AFTER (%)	−16.86% (- 18.25, - 15.36)	- 23.01% (- 24.52, - 21.23)	−21.28% (- 22.88, - 19.70)	3.32% (2.63,3.62)	- 4.03% (- 7.22, - 0.73)
	Time-invariant FE	Pair	Pair	Pair	Pair	Pair
	Time FE	None	None	None	None	None
	Number of customers	5,400	5,400	5,400	5,096	5,096
	Observations	49,291	49,291	49,291	45,284	45,284
	R^2	0.46	0.59	0.59	0.23	0.41
With monthly fixed effect	TREAT	0.45**	- 0.17	0.75	0.19***	- 0.06***
		(0.19)	(0.32)	(0.64)	(0.01)	(0.02)
	AFTER	0.61**	2.20***	3.77***	0.07***	0.03
		(0.26)	(0.46)	(0.90)	(0.01)	(0.03)
	TREAT × AFTER	- 5.56***	- 12.75***	- 23.39***	0.20***	- 0.09***
		(0.26)	(0.46)	(0.90)	(0.01)	(0.03)
	TREAT × AFTER (%)	−16.74% (- 18.32, - 15.10)	- 22.76% (- 24.33, - 21.16)	−21.00% (- 22.52, - 19.31)	3.32 % (2.64,3.62)	- 4.31 (- 7.24, - 0.64)
	Time-invariant FE	Pair	Pair	Pair	Pair	Pair
	Time FE	Month	Month	Month	Month	Month
	Number of customers	5,400	5,400	5,400	5,096	5,096
	Observations	49,291	49,291	49,291	45,284	45,284
	R^2 *(FE)*	0.46	0.59	0.59	0.42	0.24

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. (3) for *assortment size basket size*, *basket value*, *slot cost* and *proximity* accounting for monthly time fixed effects. Cluster standard errors are taken at a pair level. The percentage changes are included with non-parametric bootstrapped 95% confidence intervals (in parentheses) based on 500 random draws. Significance levels: 10% (*), 5% (**), and 1% (***).

Please cite this article as: L. Wagner, C. Pinto and P. Amorim, On the Value of Subscription Models for Online Grocery Retail, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2020.05.011

estimates. Similarly, treated members would have needed to be 8, 1.6, 3 and 1.2 times more likely to be treated before our conclusions about monthly spend, assortment size, basket size and basket value would change. In summary, our main findings on basket values are more sensitive to hidden biases than those related to monthly spend or basket size.

Customer ordering behavior over purchase trajectory. To rigorously validate the parallel trend assumption, we augment our main specification Equation (1) with leads and lags as follows:

$$\mathrm{OUTCOME}_{p,i,t} = \alpha\ \mathrm{TREAT}_{p,i} + \sum_{t=-5}^{\geq 5} \beta_t\ \mathrm{PURCHASE}_t + \sum_{t=-5}^{\geq 5} \gamma_t\ \mathrm{TREAT}_{p,i} \times \mathrm{PURCHASE}_t + \lambda_p + \epsilon_{p,i,t}, \quad (2)$$

where the dummy variable $\mathrm{PURCHASE}_t$ ranges from five orders prior to treatment to five or more purchases after treatment. To validate the existence of a parallel trend prior to treatment we require the leads γ_t for $t = \{-5, \ldots, -1\}$ to be insignificant. The results of this specification are presented in Table 7. As can be seen, the coefficients in the five leads are insignificant for most outcome variables. Note that these leads are significant for slot cost some purchases prior to treatment. We argue that, because treated customers have purchased more expensive slots prior to treatment, we at worst underestimate the treatment effect on this outcome variable. We therefore believe that reducing the sample size further would mean more drawbacks than benefits. Post-treatment, subscription consumers are significantly changing their behavior with respect to all outcome variables and purchase occasions, although the treatment effect seem to slowly fade out the more distant a purchase is from the time a customer subscribes.

Use of logarithmic scale. Many researchers in the operations research and management science community have used multiplicative models to analyze treatment effects on monetary outcome variables (Gallino, Moreno, and Stamatopoulos (2016); Gaur, Fisher, and Raman (2005)), presumably because these variables are often not normally distributed. We verify that our main results still hold when outcome variables are logged. For convenience, we provide a comparison between both our linear model and the logarithmic one in Table 8. As we can see, the estimated treatment effects under both specifications coincide in direction and are either similar in magnitudes (assortment size, basket size), or are higher (monthly spend, frequency, basket value) in the semi-log specification. Because of this, the magnitudes reported with the linear specification are at worst underestimates.

Fixed Effects Grocery retail can be greatly affected by the time at which a customer purchases. To verify that our results are robust to the purchase occasion, we include monthly fixed effects to our main model. As seen in Table 9, the treatment effect maintains its direction and magnitude. Thus, estimating the model with or without monthly fixed effect plays a relatively minor role.

Seemingly unrelated regression If a customer orders less frequently, one may expect that she also spends less per month at the retailer's online channel. This implies that frequency and monthly spend are expected to be correlated. With the objective of explaining consumers' ordering behaviour as a whole, correlations among dependent variables should be accounted for. Seemingly Unrelated Regression (SUR) is a method of doing so. Instead of estimating each DID equation separately– as we have done thus far– the SUR method considers these equations simultaneously. Table 10 and Table 11 confirm our intuition that outcome variables are correlated, justifying the use of an SUR method.

We run simultaneous regressions for the two point estimates (frequency and monthly spend) separately from the ones that are specified for each purchase (assortment size, basket size, basket

Table 10
The correlations of the residuals

	Monthly spend	Frequency
Monthly spend	1.000	0.683
Frequency	0.683	1.000

Table 11
The correlations of the residuals

	Assortment size	Basket size	Basket value
Assortment size	1.000	0.733	0.695
Basket size	0.733	1.000	0.838
Basket value	0.695	0.838	1.000

Table 12
Seemingly unrelated regression: The effect of subscription on consumer ordering behavior

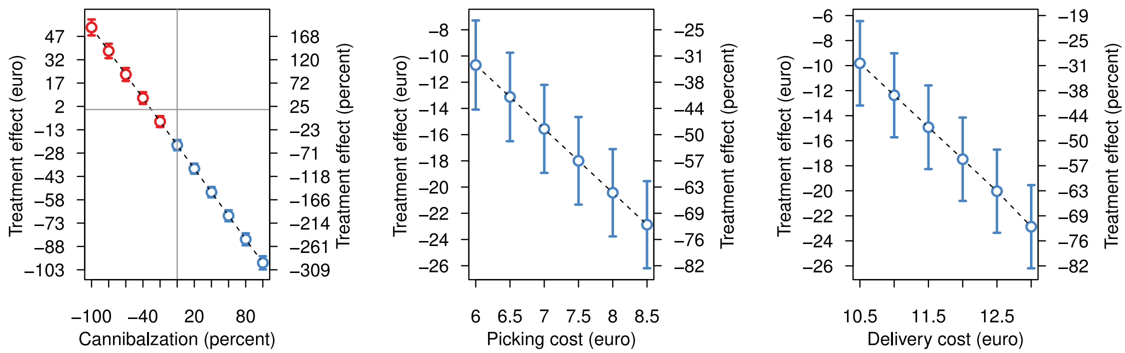
	Dependent variable	
	Monthly spend (€ /month)	Frequency (purchases/month)
TREAT	0.26	0.00
	(2.60)	(0.02)
AFTER	- 14.10***	- 0.21***
	(2.60)	(0.02)
TREAT × AFTER	107.00***	1.56***
	(3.67)	(0.04)
TREAT × AFTER (%)	85.2%	131.6%
	(78.95,91.51)	(125.3,138.4)
Time-invariant FE	Pair	Pair
Time FE	None	None
Number of customers	5,400	5,400
Observations	10,800	10,800
R^2	0.68	0.65

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. (1) for *monthly spend* and *frequency* simultaneously. The percentage changes are included with non-parametric bootstrapped 95% confidence intervals (in parentheses) based on 500 random draws. Significance levels: 10% (*), 5% (**), and 1% (***).

Table 13
Seemingly unrelated regression: The effect of subscription on consumer ordering behavior

	Dependent variable		
	Assortment size (units/purchase)	Basket size (units/purchase)	Basket value (€ /purchase)
TREAT	1.16*	- 0.08	1.05
	(0.39)	(0.64)	(1.26)
AFTER	0.65	2.25*	5.58***
	(0.39)	(0.64)	(1.26)
TREAT × AFTER	- 6.99***	- 14.79***	- 26.32***
	(0.55)	(0.91)	(1.78)
TREAT × AFTER (%)	−20.38%	−25.64%	−23.19%
	(- 24.16, - 17.06)	(- 28.91, - 22.89)	(- 26.57, - 20.37)
Time-invariant FE	Pair	Pair	Pair
Time FE	None	None	None
Number of customers	5,400	5,400	5,400
Observations	49,291	49,291	49,291
R^2	0.58	0.71	0.71

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. (1) for *assortment size*, *basket size* and *basket value* simultaneously. The percentage changes are included with non-parametric bootstrapped 95% confidence intervals (in parentheses) based on 100 random draws. Significance levels: 10% (*), 5% (**), and 1% (***).



(a) Sensitivity to cannibalization (b) Sensitivity to picking cost (c) Sensitivity to delivery cost

Fig. 2. Sensitivity analysis with respect to various parameters

Table 14
Parameters for Profit baseline model

Parameter name	Description	Value
margin	retailer's average margin on sold items	20 %
Picking	cost of collecting the items	8.50€
Delivery(Slot Width)=$m \times$ SlotWidth $+ b$		
m	delivery cost slope	13.055€
b	delivery cost constant	-0.5785€

Note: The table reports the parameters used to compute PROFIT for our base-line model. The parameters of Delivery(Slot Width) are based on Table 2 of Boyer et al. (2009). Here we fitted a linear trend to their data. This results in a slot width-dependent delivery cost of 7.27€ -11.898€ for a slot width range of 2.5 to 10 hours.

value). As we can see in the Tables 12 and 13, there was no difference in the effect of subscription on (shopping) frequency and monthly spend, whether we estimated the DID with SUR or without. This is no longer true when using assortment size, basket size and basket value as dependent variables: the effect of subscription on these covariates is slightly stronger when SUR is employed.

7. Impact of subscriptions on retailer's profit

For each customer in a pair p, we let T and A abbreviate *TREAT* and *AFTER*, respectively. Moreover, we define $\mathcal{P}_{T,A}$ as the sets of all purchase incidences prior and after treatment. Thus the four sets $\mathcal{P}_{T,A}$ per customer pair, consists of all purchase incidences during the respective period. Then for each customer we calculate the profit as follows:

For all $T = 0$ or $A = 0$:

$$\mathrm{PROFIT}_{T,A} = \sum_{j\in\mathcal{P}_{T,A}} \underbrace{\mathrm{margin} \times \mathrm{BaskV}_{j,T,A}}_{\text{revenue from core business}} - \underbrace{\mathrm{Picking} - \mathrm{Delivery(SlotWidth)}_{j,T,A}}_{\text{operational cost}} + \underbrace{\mathrm{SlotCost}_{j,T,A}}_{\text{revenue from delivery service}}$$

Otherwise, $T = 1$ and $A = 1$:

$$\mathrm{PROFIT}_{T=1,A=1} = \underbrace{\mathrm{margin}\Big(\sum_{j\in\mathcal{P}_{T,A}} \mathrm{BaskV}_{j,T,A} - \max\Big\{\sum_{j\in\mathcal{P}_{T,A}} \mathrm{BaskV}_{j,T=1,A=1} - \sum_{j\in\mathcal{P}_{T,A}} \mathrm{BaskV}_{j,T=0,A=1}, 0\Big\} \times c\Big)}_{\text{revenue from core business}}$$
$$- \sum_{j\in\mathcal{P}_{T,A}} \underbrace{\mathrm{Picking} - \mathrm{Delivery(SlotWidth)}_{j,T,A}}_{\text{operational cost}} + \underbrace{\mathrm{SubscPrice}_{j=1,T,A}}_{\text{revenue from delivery service}}$$

The *revenue from core business* encompasses the total *basket value* (BaskV) multiplied by retailer's profit *margin*, when customers pay delivery on a per-order basis (T=0 or A=0). Once a customer signs up to the subscription plan, the profit equation also accounts for possible cannibalization of offline sales in the retailer's stores. In particular, we calculate how much more a treated customer has spent than a control customer after treatment, and define the cannibalization level c in percentage. The *operational cost* entails the cost of collecting the items a customer demands at one of the retailer's physical stores, and the cost of last-mile delivery as a function of the slot size selected, denoted with *Picking* and *Delivery(Slot Width)*$_{j,T,A}$, respectively. The latter probably merits some further explanation. In line with previous research (Boyer, Prud'homme, & Chung, 2009), we assume that narrower delivery slots are more expensive for retailers. For the sake of simplicity, its functional form is assumed to be linear, i.e., *Delivery(Slot Width)*$_{j,T,A} = m \times \mathrm{SlotWidth} + b$.[14] Finally, *revenue from delivery service* is composed of either, the price a consumer pays for the slot, denoted with *SlotCost*$_{j,T,A}$ under the pay-per-order scheme, or the one-time subscription price (*SubscPrice*$_{j=1,T=1,A=1}$) paid to qualify for the free delivery service. We gather values missing in our dataset from two sources: (i) the focal retailer provided us with values for *margin* and *Picking*, and (ii) previous research (Boyer et al., 2009) helped us to derive delivery cost data for each slot (*Delivery(Slot Width)*$_{j,T,A}$).

Boyer et al. (2009) collected data about customer density and delivery window length from interviews with managers in the e-commerce industry. Their findings suggest that offering a 3 hour delivery window is 3045% more expensive than offering unattended (9 hour delivery window) delivery. We use the parameters in their paper to derive delivery cost data for each slot (*Delivery(Slot Width)*$_{j,p_{T,A}}$). In the absence of detailed routing data from our partner, and due to the similarity of delivery window length (2.5 to 10 hours), we have settled for using these findings as our basis for constructing the slot width-dependent delivery costs, *Delivery(Slot Width)*$_{j,p_{T,A}}$. The specific parameters used for our baseline model are given in Table 14.

We obtain estimates of the treatment effect on the outcome variable *PROFIT* by employing the DID Eq. (1) for various cannibalisation parameters. The results are presented in Panel (a) of

[14] The delivery cost is zero when the products are picked up by the customer.

Figure 2 and Tables 20 and 21 in the Appendix. As can be seen, once customers subscribe, they become, on average, less profitable than the pay-per-usage customer, notwithstanding the assumption about the shift in demand from the retailer's offline business. The total losses to the retailer's profit range from approximately 23€ (−69.26%) to 99€ (- 306.2%) per customer, depending on how many of the online channel's subscription customers were previously loyal to the retailer's offline channel (i.e, $c \geq 0$). This conclusion, however, would not necessarily hold if a positive spillover effect (i.e, $c < 0$) from subscription on a retailer's physical store were to take precedence over a negative one. Were there a positive spillover effect above ≈30%, a percentage which in this context is fairly high, our conclusion would be reverted. We can thus safely conclude that subscription has a negative impact on a grocery retailer's profit.

However, one may reasonably be concerned that the 904 customers[15] who had to be dropped because of our matching method[16] are the ones responsible for an important share of the retailer's profit. Excluding them would therefore bias our results. Table 22 displays the profits and their relative changes prior to and after treatment for the treatment and the control samples, as well as for the overall treatment group. The results suggest that the overall treatment group behaves similarly, in relative terms, to the treated sub-sample used in this analysis, and that the profit loss for the overall treatment group is higher.[17] In other words, the results displayed in this section are, at worst, underestimations.

These conclusions, of course, depend on parameters we did not directly obtain from our dataset. We next re-estimate the impact of subscription on profits, reducing parameters of *Picking* and $DeliveryCost(Slotwidth)_{j,\mathrm{T,A}}$ (parameter b) by up to approximately 20% while holding the cannibalization percentage at 0%, depicted in Panel (b) and (c) of Fig. 2 and Tables 23 and 24.[18]

But how much would these costs have to be adjusted in order to lead us to the opposite conclusion? To answer this question, we successively reduced the two parameters until we obtained a significantly positive treatment effect. In the absence of cannibalization, we find that the retailer would have had to drop the picking cost by at least 64.71% and the delivery cost (parameter b) by 38.72%–numbers which are fairly difficult to drop down to in practice.

8. Data-driven customer selection

Up to this point, we have provided evidence that, on average, subscription plans reduce retailers' profits. We now go on to ask what would happen if the retailer could decide who is able to see offers and promotions for such subscription plans. By going from advertising the plan to *all* customers to selecting just a *few*, the retailer could turn an otherwise unprofitable business into a profitable one. In this section, then, we develop data-driven algorithms to predict whether allowing a customer to buy into the subscription plan results in higher or lower profits. The algorithm learns from only the *first purchase incidence* available in the dataset[19] to inform this prediction. The development of the algorithm follows three main steps: First, variables with sufficient predictive power are selected. Secondly, the model is trained and predictions are made, and finally, the prediction model is evaluated.

Table 15
Predictive power of covariates

	Dependent variable
	incremental profit after subscription
Constant	- 49.856***
	(5.360)
Assortment	- 0.459***
size	(0.120)
Basket	- 0.026
size	(0.082)
Basket	0.342***
value	(0.038)
Slot	1.206
cost	(0.840)
Number of customers	2,025
Observations	2,025
R^2	0.082

Note: The table reports the estimated coefficients and standard errors (in parenthesis) for several covariates to explain the incremental profit gain after treatment on the training set. The data-driven subscription targeting section is developed for the base-line model for cannibalization of zero and the parameters given in Table 14.

Variable selection. Here we split the data into a 75% training and a 25% test set and run a linear regression model to decide which covariates are sufficiently informative.[2] Table 15 reports the outcome on the training set. As can be seen, only the *assortment size* and *basket value* of a customer's first purchase are strong (significant) enough to predict the difference in profits between a subscription customer and her control counterpart. These are the ones we kept.

Training and Prediction. In the next phase, we used a linear regression model[21] to train the 75% of data and predict the outcome for the remaining 25%. We resampled the data set 5000 times and repeated the training and prediction procedure.

Evaluation. Any predicted profit loss would suggest that the customer should not be targeted, while any profit gain would suggest the opposite. Table 16 cross-tabulates the number of positive and negative profit differences for both our prediction model and an oracle[22] for the 5000 iterations. As one can see, the oracle would not have allowed 63% of the 675 customers in the test sample to sign up for the subscription plan, while our prediction model would have excluded 91% of all customers.

We also calculate how limiting the pool of customers who can sign up for the subscription compares to the status quo in terms of profit gains. For example, if a customer generates 30€ less than her control counterpart and the algorithm correctly predicts that this customer should not be targeted, then it has saved the retailer 30€. By contrast, in cases where the algorithm predicts a positive profit, no money would be saved or lost. In short, only if our algorithm reverts the decision to advertise the subscription plan to a customer does the retailer save or lose money. Using the same logic, we additionally show how an oracle would have performed. Fig. 3 depicts the average profit gain obtained across the 5000 rounds for both the algorithm and the oracle. By restricting access to the subscription plan, our model is able to improve profit per-customer

[15] 904=3604-2700 customers (Overall Treatment group-Treatment group), please see Table 3.

[16] This is because, even though these customers signed up for treatment no sufficiently close control group could be found

[17] The 904 customers have shopped more often per month (3.04 times/month) than the treated customer in our final sample (1.19 times/month) prior to treatment. After treatment, these 904 customers did not substantially increase their shopping frequency (3.09 times/month).

[18] We chose 0% as the base-line as this represents the "best case scenario".

[19] We experimented with the length of the learning period. For instance, we have extended this period to the first two purchase incidences prior to treatment. The accuracy of the results is very similar, but we would have needed to reduce the number of customers to 1821 because the remaining ones had already purchased the subscription plan after the first purchase.

[20] One may expect that some of the covariates are correlated. We addressed this issue by running the regression equation multiple times, each time removing some covariates. The results remained fairly similar.

[21] We also tested more complex models such as random-forests and logit models. As their predictive powers were comparable, we here present the outcome of the simpler linear regression model.

[22] The oracle is an algorithm that uses the exact post-profit differences between the treated and control.

Table 16
Crosstable of predicted versus actual incremental profit sign

		Predicted			
			Negative Profit Mean (Std.Dev)	*Postive Profit* Mean (Std.Dev)	Observations
Actual (Oracle)	*Positive Profit*	Mean (Std.Dev)	31.71% (1.84)	5.08% (0.95)	248
	Negative Profit	Mean (Std.Dev)	59.12% (1.65)	4.09% (1.07)	427
		Observations	613	62	675

The table display the sign of the incremental profit after treatment for the test set over 5000 random draws.

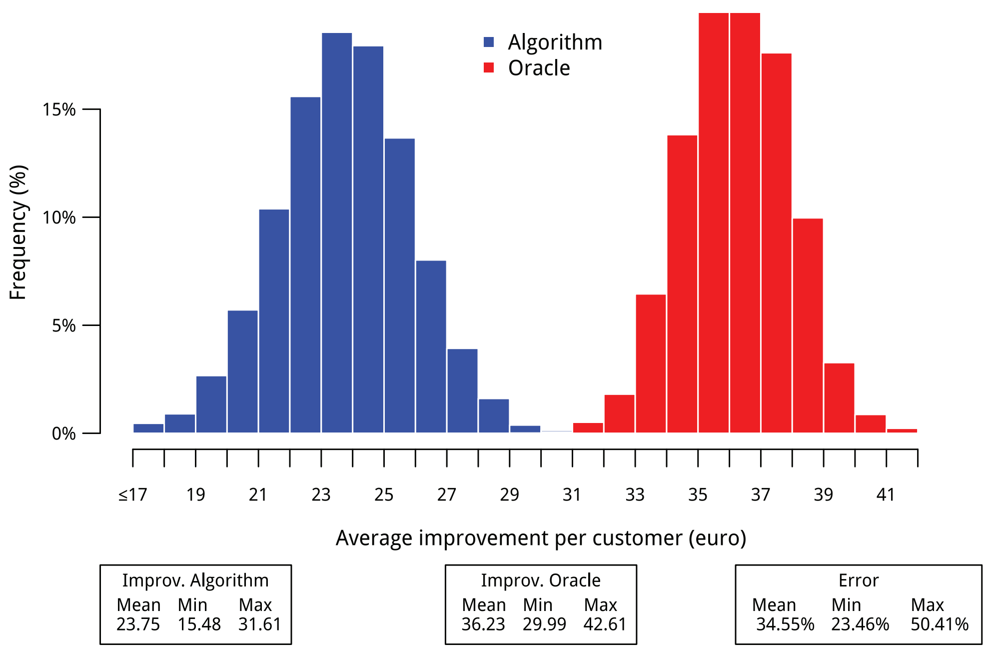


Fig. 3. Results from Algorithm and Oracle, across the 5000 repetitions.

by, on average, 23.75€ with a margin of error of 34.55%.[23] Interestingly, none of the 5000 rounds resulted in a loss for the company- A great improvement to the results obtained in Section 7.

9. Extension: Adding a threshold to the subscription service

So far, we have studied the impact of subscriptions for free deliveries from a marketing and operational perspective. Our analysis has revealed that the adoption of a subscription-based delivery service leads to increasing consumer purchase frequency, while reducing assortment and basket size. In this extension, we exploit the fact that the focal retailer offered two distinct subscription plans during the time of our study. This situation allows us to analyse whether changing key features of the plan impacts customers' interaction with the retailer's online channel.

Research hypothesis

Home delivery service payment plans come in all shapes and sizes. For instance, Amazon offers *unlimited* home deliveries through their grocery service (Amazon Fresh) at zero cost to customers who already pay $14.99 per month for their Prime membership (Amazon, 2019). Conversely, Walmart does not offer any subscription scheme, with customers bearing a fixed shipping cost to have their purchases delivered to their home. Somewhere in between is what we refer to as a *contingent* plan, where delivery at no extra cost is contingent upon a minimum spend. The retailer we partnered with added such a threshold to their pre-existing subscription plan during the observation period, leaving all other contract parameters untouched (details will follow). This adjustment allows us to investigate whether consumer ordering behavior changes when this type of policy is implemented. More specifically, it begs the question whether the addition of a contingent threshold for delivery raises consumer spending, as it does when not bundled with a subscription offer. In line with Chen and Ngwe (2018), we conjecture that consumers who would normally be reluctant to increase their basket value may now opt to spend enough to at least match the threshold and avoid having to pay the delivery cost.

Hypothesis 4 (Contingency plan: Basket value)**.** Customers with a contingent plan spend more per online purchase compared to those with an unlimited subscription.

If this hypothesis is confirmed, what exactly are these consumers spending their extra money on? In particular, do they purchase more goods, or fewer, more expensive ones? How does this affect the frequency of their orders? We build our hypothesis around three possible sales mechanisms, which could lead to alterations in consumer behavior as they try to meet the contingent threshold.

- *Upselling* encourages customers to spend more by buying costlier products (usually premium products or brands), meaning that a consumer's per-session assortment and basket size remain unchanged when a contingent threshold is added to the existing subscription plan. On top of this, spending more money

[23] The margin of error was calculated as (Improvement Oracle-Improvement Algorithm)/(Improvement Oracle) for each of the 5000 rounds.

Table 17
Subscription plans offered by our partner during the time of our study

Subscription Model	Cost for the customer	Duration of the Subscription	Minimum basket value	Time period in the dataset
Unlimited	26,90€	100 days	0€	Oct. 2016 - May 2017
Contingent	26,90€	100 days	35€	Feb. 2017 - Sep. 2017

Table 18
Summary statistics after matching for the two subscriptions

Variable	*Treatment group*		*Control group*		Observations	Std. Mean differences (smd)
	Mean	Std.Dev	Mean	Std.Dev		
Monthly spend* (€)	101.838	139.258	102.249	124.558	698	0.003
Frequency (units)	0.871	0.980	0.892	0.955	698	0.022
Average assortment size* (units)	33.730	16.167	34.497	18.155	698	0.045
Average basket size (units)	58.082	34.796	58.652	37.923	698	0.016
Average basket value (€)	117.474	75.075	119.372	74.833	698	0.025
home delivery (percent)	93.822	22.060	95.601	19.810	698	0.085
Average slot cost (€)	5.021	2.197	5.151	2.011	698	0.062

Note: The table presents the observable variables used in our matching procedure to obtain the closest pairs prior to the treatment event. The delivery cost for pick-up-in-store transactions is defined as zero. The standard mean differences are pooled and presented in absolute values.*We additionally display the summary statistic of *monthly spend* and *assortment size*, even though these variables are not explicitly used as a matching variable.

on the same amount of products does not alter the frequency of shopping trips. We expect changes in the order frequency, assortment, and basket size to be insignificant under the influence of this mechanism.

- *Cross-selling* encourages customers to spend more by purchasing additional distinct (often non-essential) products. This leads us to hypothesize that consumers' assortment and basket size increase when a contingent threshold is part of the subscription plan. Because the extra spending was only performed to meet the threshold, we expect that the order frequency will be no different to that of an unlimited subscription customer.
- *Pantry loading* is when consumers increase the number of units from goods they habitually buy to both cross the threshold and increase the length of intervals between their shopping trips. Therefore, we conjecture that the order frequency decreases and the basket size increases in this scenario. As customers stick to their initial shopping intentions, we expect that changes in assortment size will be insignificant.

As a result, we formally hypothesize the following:

Hypothesis 5 (Contingency plan-mechanisms)**.**

(i) If there is evidence of an upselling mechanism, the addition of a threshold does not lead to a change in order frequency, assortment or basket size.
(ii) If there is evidence of a cross-selling mechanism, the addition of a threshold increases the assortment and basket size, while the order frequency remains the same.
(iii) If there is evidence of a pantry loading mechanism, the addition of a threshold decreases order frequency and increases basket size, while assortment size does not change.

Institutional set up

During the time of our study (October 2016 to September 2017), the retailer offered two distinct subscription plans (solely) on their website. Originally, customers were offered unlimited deliveries regardless of the total value of their purchases. We refer to this subscription plan as the *unlimited* subscription model. Rolled out (region-by-region) during the last few months of the unlimited model's existence, the retailer's new plan continued to offer 100 days' unlimited delivery at the same price, but with each delivery now contingent upon a minimum spend of 35€. We call this the *contingent* model.[24]

The features of both plans are summarized in Table 17.

Such a setup is ideal for empirical identification, because both the subscription plan's price and period of validity remain the same, with the only observable change being the introduction of a minimum order value of 35€.

Model and Results

To validate the research hypothesis, we make use of an identification strategy based on combined-matching and DID, similar to the one explained in Section 5. We deviate from the original identification in the following way: because we are interested in measuring the effect of the explanatory variable (the addition of the contingent threshold to the subscription), we pair customers who were offered the *contingent* plan with those who purchased the *unlimited* one. As before, our goal is to ensure that paired customers exhibit "similar" patterns prior to their signing-up for the plan.[25] Contrary to the matching procedure described previously, we use a nearest-neighbour matching method for *all* matching characteristics (order *frequency*, average *basket size*, average *basket value*, percentage of *home delivery* and average delivery *slot cost*) to increase the number of possible matches. The statistics of the resulting 698 pairs can be found in Table 18. Similar to the specification in Eq. (1), we measure the treatment effect using the following equation:

$$\begin{aligned}\text{OUTCOME}_{p,i,t} = {} & \alpha\text{TREAT}^C_{p,i} + \beta\text{AFTER}_t + \gamma\text{TREAT}^C_{p,i} \\ & \times \text{AFTER}_t + \lambda_p + \epsilon_{p,i,t},\end{aligned} \tag{3}$$

where the dummy variable $\text{TREAT}^C_{p,i} = 1$, if a customer in pair (p) purchased a contingent subscription plan, with $\text{TREAT}^C_{p,i} = 0$ indicating that the customer subscribed to the unlimited plan. The results of the analysis are depicted in Table 19.

[24] Customers seeing publicity for the subscription plan on the company's website would see a clause specifying that subscribers in their region were subject to a 35€ minimum spend in order to get free delivery. In such areas, customers were not able to buy the unlimited plan any more. That is, there would be regions in which some customers still had an unlimited subscription, while more recent sign-ups would be limited to the contingent plan.

[25] In this case, each customer has a well-defined time period both prior to and after the signing of the plan, even if time periods do not necessarily coincide.

Table 19
Customer ordering behavior - Contingent vs Unlimited plans

	Dependent variable					
	Monthly spend (€ /month)	Frequency (purchases/month)	Assortment size (units/purchase)	Basket size (units/purchase)	Basket value (€ /purchase)	Average Profit (€ /purchase)
$TREAT^C$	- 0.41	- 0.02	- 0.48	0.29	1.64	- 0.17
	(7.85)	(0.07)	(0.39)	(0.87)	(1.79)	(0.63)
AFTER	113.88***	1.59***	- 5.60***	- 11.93***	- 25.74***	15.44***
	(7.85)	(0.07)	(0.40)	(0.89)	(1.82)	(0.63)
$TREAT^C$	11.55	0.02	2.02***	3.61***	7.56***	1.69*
× AFTER	(11.11)	(0.09)	(0.51)	(1.14)	(2.33)	(0.89)
TREAT × AFTER	11.79%	7.02%	6.47%	5.93%	6.72%	19.21%
(%)	(- 20.79,47.50)	(- 29.06,43.46)	(2.75,10.51)	(2.75,8.80)	(3.24,10.01)	(- 19.56,53.38)
Time-invariant FE	Pair	Pair	Pair	Pair	Pair	Pair
Time FE	None	None	None	None	None	None
Number of customers	1,396	1,396	1,396	1,396	1,396	1,396
Observations	2,792	2,792	16,688	16,688	16,688	2,792
R^2	0.51	0.56	0.31	0.35	0.36	0.60

Note: The table reports the estimated coefficient values and standard errors (in parentheses) of Eq. 3 for *monthly spend*, *frequency*, *assortment size*,*basket size* and *basket value*. The first two regressions make use of a two period (prior and after) difference-in-difference models, where the average is calculated as variable/period length (i.e., each customer represents two observations, while the unit of analysis for assortment size, basket size and basket value is a purchase. The *Average Profit* denotes the average profit per purchase and customer and is calculated for the baseline model with the following equation:
$\text{Average Profit}_{T,A} = margin \times \overline{BaskV}_{T,A} - Picking - \overline{Delivery}(\text{SlotWidth})_{T,A} + \overline{SlotCost}_{T,A} + SubscPrice_{T,A}$, where $\bar{x}$, denotes the average. Cluster standard errors are taken at a pair level. Significance levels: 10% (*), 5% (**), and 1% (***).

The findings show a positive and significant treatment effect on basket value: consumers spend 7.56€ more per purchase when presented with a contingent threshold – a 6.4% increase. Moreover, we find that consumers increase their assortment and basket size rather than opting to buy more expensive versions of their normal groceries (upselling) when there is a contingent threshold. Also notice that consumers do not alter the frequency of their visits, resulting in an average profit increase of 1.69€ per purchase. This leads us to conclude that compared to customers with an *unlimited* plan, customer subject to a *contingent* one are more vulnerable to cross-selling than the other sales mechanisms.

10. Conclusions

We show that, although subscription-based delivery services achieve their marketing goals – i.e., increasing revenue and market share, they also cause significant logistical headaches for the retailer, with customers becoming more demanding in terms of their delivery preferences. That is, customers with a subscription increase their use of the home delivery service and choose narrower slots, mainly in the morning and at night, thereby increasing last-mile delivery costs. These operational challenges are especially difficult to overcome in the online grocery retail business, where margins are slim and competition is high.

We show that the revenue obtained from subscription customers are canceled out by the extra cost of delivery. To mitigate against these losses, we develop a data-driven algorithm which is aimed at assisting retailers in selecting only those customers who will generate positive incremental profit. We obtained these results using data from the largest omni-channel grocery retailer in the focal country, who sell a huge variety of products both online and offline. A setting which can be expected to hold for any typical large grocery retail setting.

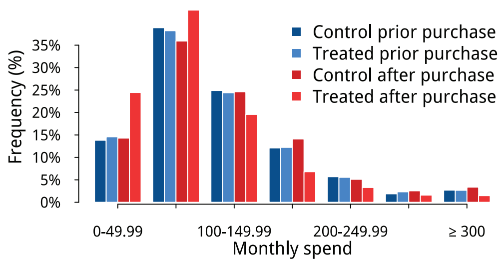
Furthermore, we estimate the causal impact of introducing a minimum-spend threshold into the subscription plan. Our findings indicate that customers subsequently buy more items (assortment size, basket size) and spend more (basket value) per purchase, being now more susceptible to cross-selling, rather than upselling or pantry loading.

Our results thus show that companies wanting to offer subscription plans with the goal of both expanding market share and making a profit need to be cautious from the outset in setting certain parameters. Moreover, our results also emphasize the need for dynamic slot pricing methods and other approaches that nudge customers towards the selection of some slots instead of others. For example, Amazon offers low-value vouchers to users who agree to wait longer for their goods. Putting such policies into practice may help in lowering the last-mile delivery costs at a low price. We are also at pains to point out that retailers should take subscription customers into account when deciding on a delivery-slot pricing strategy. How these parameters can be tweaked to further benefit retailers, is an intriguing avenue for future research.

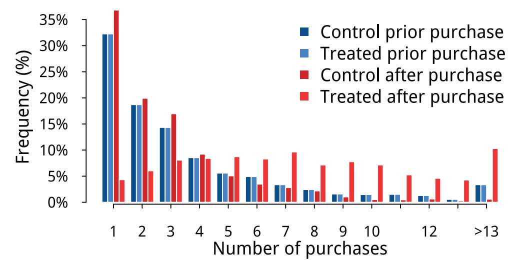
Acknowledgements

The authors would like to thank two anonymous referees for their helpful comments and insights. Catarina Pinto acknowledges FCT for her PhD grant (PD/BD/128468/2017).

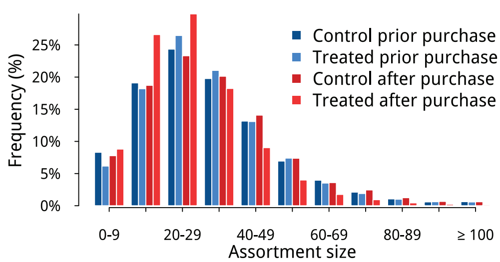
Appendix



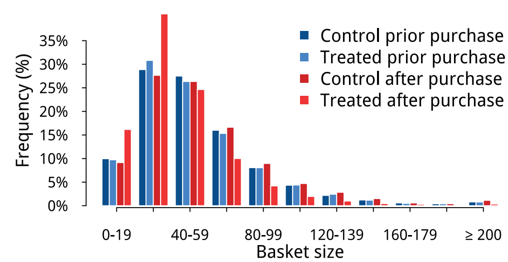
(a) Distribution of *monthly spend* prior and after treatment



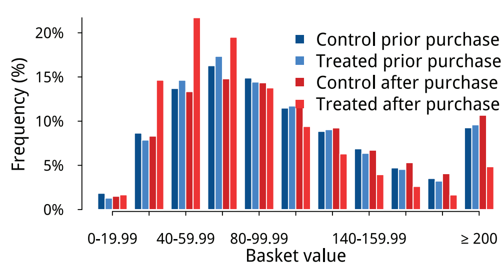
(b) Distribution of *number of purchases* prior and after treatment



(c) Distribution of *assortment size* prior and after treatment



(d) Distribution of *basket size* prior and after treatment



(e) Distribution of *basket value* prior and after treatment

Fig. 4. Distribution of several outcome variables prior and after treatment

Table 20
The effect of subscription on retailer's profit: as a function of cannibalisation c (1).

	Dependent variable				
	Profit with cannibalization (c) of				
	- 100%	- 80%	- 60%	- 40%	- 20%
TREAT	0.913	0.913	0.913	0.913	0.913
	(1.842)	(1.665)	(1.506)	(1.370)	(1.265)
AFTER	- 80.005***	- 64.875***	- 49.745***	- 34.615***	- 19.485***
	(1.199)	(1.178)	(1.206)	(1.279)	(1.389)
TREAT × AFTER	52.782***	37.651***	22.521***	7.391***	- 7.739***
	(2.605)	(2.355)	(2.130)	(1.937)	(1.788)
TREAT × AFTER	167.68%	120.3%	72.91%	25.52%	-21.87%
(%)	(150.73,188.06)	(106.03,136.06)	(61.33,86.43)	(14.08,37.63)	(- 30.97, - 10.91)
Time-invariant FE	Pair	Pair	Pair	Pair	Pair
Time FE	None	None	None	None	None
Number of customers	5,400	5,400	5,400	5,400	5,400
Observations	10,800	10,800	10,800	10,800	10,800
R^2	0.546	0.569	0.597	0.629	0.661

Note: The table reports the estimated coefficient values and standard errors (in parentheses) for the *PROFIT* with varying cannibalization assumptions. Significance levels: 10% (*), 5% (**), and 1% (***).

Table 21
The effect of subscription on retailer's profit: as a function of cannibalisation c (2).

	Dependent variable					
	Profit with cannibalization (c) of					
	0%	20%	40%	60%	80%	100%
TREAT	0.913	0.913	0.913	0.913	0.913	0.913
	(1.199)	(1.178)	(1.206)	(1.279)	(1.389)	(1.530)
AFTER	- 4.354***	10.776***	25.906***	41.036***	56.166***	71.296***
	(1.199)	(1.178)	(1.206)	(1.279)	(1.389)	(1.530)
TREAT × AFTER	- 22.869***	- 37.999***	- 53.130***	- 68.260***	- 83.390***	- 98.520***
	(1.695)	(1.666)	(1.706)	(1.809)	(1.965)	(2.163)
TREAT × AFTER	−69.26%	- 116.64%	−164.03%	- 211.42%	- 258.81%	- 306.2%
(%)	(- 79.45, - 58.37)	(- 127.18, - 106.39)	(- 177.91, - 150.92)	(- 225.94, - 198.67)	(- 278.97, - 239.50)	(- 330.63, - 283.69)
Time-invariant FE	Pair	Pair	Pair	Pair	Pair	Pair
Time FE	None	None	None	None	None	None
Number of customers	5,400	5,400	5,400	5,400	5,400	5,400
Observations	10,800	10,800	10,800	10,800	10,800	10,800
R^2	0.689	0.708	0.717	0.716	0.708	0.696

Note: The table reports the estimated coefficient values and standard errors (in parentheses) for the *PROFIT* with varying cannibalization assumption. Significance levels: 10% (*), 5% (**), and 1% (***).

Table 22
Summary statistics of profit components for different groups

Variable	*Prior Treatment*		*After Treatment*			
	Mean	Std.Dev	Mean	Std.Dev	Rel. difference	Observations
Treatment group	32.841	65.631	5.618	85.276	- 82.89%	2700
Control group	31.928	65.316	27.574	48.765	- 13.64%	2700
Overall treatment group	42.738	124.552	6.759	100.794	−84.18%	3,604

Note: The table presents the profit components prior and after treatment for the treatment group, the control group, and the overall treatment group.

Table 23
The effect of subscription on profit generation (2)

	Dependent variable					
	Profit with picking cost of					
	6	6.5	7	7.5	8	8.5
TREAT	0.913	0.913	0.913	0.913	0.913	0.913
	(1.226)	(1.217)	(1.211)	(1.205)	(1.201)	(1.199)
AFTER	- 6.662***	- 6.200***	- 5.739***	- 5.277***	- 4.816***	- 4.354***
	(1.226)	(1.217)	(1.211)	(1.205)	(1.201)	(1.199)
	(1.199)	(1.178)	(1.206)	(1.279)	(1.389)	(1.530)
TREAT × AFTER	- 10.689***	- 13.125***	- 15.561***	- 17.997***	- 20.433***	- 22.869***
	(1.734)	(1.722)	(1.712)	(1.704)	(1.699)	(1.695)
TREAT × AFTER	−24.89%	- 32.09%	−40.00%	- 48.74%	- 58.43%	- 69.26%
(%)	(- 32.66, - 17.00)	(- -39.75, - 24.38)	(- 48.14, - 31.82)	(- 56.89, - 39.84)	(- 67.72, - 48.82)	(- 79.48, - 58.46)
Time-invariant FE	Pair	Pair	Pair	Pair	Pair	Pair
Time FE	None	None	None	None	None	None
Number of customers	5,400	5,400	5,400	5,400	5,400	5,400
Observations	10,800	10,800	10,800	10,800	10,800	10,800
R^2	0.686	0.687	0.688	0.688	0.689	0.689

Note: The table reports the estimated coefficient values and standard errors (in parentheses) for the *PROFIT* with varying cannibalization assumption. Significance levels: 10% (*), 5% (**), and 1% (***).

Table 24
The effect of subscription on retailer's profit: as a function of delivery cost (constant b).

	Dependent variable					
	Profit with delivery cost (constant b) of					
	10.5	11	11.5	12	12.5	13.055
TREAT	0.926	0.924	0.921	0.918	0.916	0.913
	(1.217)	(1.210)	(1.205)	(1.202)	(1.199)	(1.199)
AFTER	- 6.472***	- 6.058***	- 5.643***	- 5.229***	- 4.815***	- 4.354***
	(1.217)	(1.210)	(1.205)	(1.202)	(1.199)	(1.199)
TREAT × AFTER	- 9.815***	- 12.370***	- 14.924***	- 17.479***	- 20.034***	- 22.869***
	(1.721)	(1.712)	(1.705)	(1.699)	(1.696)	(1.695)
TREAT × AFTER	−23.21%	- 30.62 %	38.7%	- 47.55%	- 57.27%	- 69.26 %
(%)	(- 32.00, - 14.22)	(- 38.93, - 22.11)	(- 47.07, - 30.19)	(- 56.17, - 38.00)	(- 67.13, - 47.80)	(- 79.43, - 59.85)
Time-invariant FE	Pair	Pair	Pair	Pair	Pair	Pair
Time FE	None	None	None	None	None	None
Number of customers	5,400	5,400	5,400	5,400	5,400	5,400
Observations	10,800	10,800	10,800	10,800	10,800	10,800
R^2	0.690	0.690	0.690	0.690	0.690	0.689

Note: The table reports the estimated coefficient values and standard errors (in parentheses) for the *PROFIT* with varying cannibalization assumption. Significance levels: 10% (*), 5% (**), and 1% (***).

Table 25
Sensitivity Analysis unobservable variables

	Monthly spend			*Frequency*			*Mean assortment size*			*Mean basket size*			*Mean basket value*		
Γ	U. Bound P-Value	L. Bound HL Est.	U. Bound HL Est.	U. Bound P-Value	L. Bound HL Est.	U. Bound HL Est.	U. Bound P-Value	L. Bound HL Est.	U. Bound HL Est.	U. Bound P-Value	L. Bound HL Est.	U. Bound HL Est.	U. Bound P-Value	L. Bound HL Est.	U. Bound HL Est.
1	< 0.0001	99.32	99.32	< 0.0001	1.67	1.67	< 0.0001	-4.42	-4.42	< 0.0001	-10.46	-10.46	< 0.0001	-0.21	-0.21
1.2	< 0.0001	89.62	109.12	< 0.0001	1.27	1.77	< 0.0001	-5.82	-3.02	< 0.0001	-12.56	-8.46	0.0505	-0.41	-0.01
1.4	< 0.0001	81.72	117.62	< 0.0001	1.17	1.77	< 0.0001	-7.02	-1.82	< 0.0001	-14.36	-6.76			
1.6	< 0.0001	74.92	125.02	< 0.0001	1.17	1.87	0.0028	-8.02	-0.92	< 0.0001	-15.96	-5.36			
1.8	< 0.0001	69.02	131.62	< 0.0001	0.97	1.87	0.4087	-9.02	-0.02	< 0.0001	-17.36	-4.16			
2	< 0.0001	63.82	137.62	< 0.0001	0.97	2.07				< 0.0001	-18.76	-3.16			
3	< 0.0001	44.52	161.02	< 0.0001	0.77	2.17				0.9216	-24.06	0.71			
4	< 0.0001	31.32	178.02	< 0.0001	0.67	2.47									
5	< 0.0001	21.22	191.42	< 0.0001	0.57	2.67									
6	< 0.0001	13.12	202.52	< 0.0001	0.37	2.67									
8	0.4467	0.42	220.22	< 0.0001	0.27	2.97									
10				< 0.0001	0.27	3.07									
12				< 0.0001	0.07	3.07									
14				< 0.0001	0.07	3.27									
16				0.0001	0.07	3.27									
18				0.0037	-0.03	3.37									
20				0.0366	-0.03	3.37									
22				0.1603	-0.03	3.57									

Note: Γ is odds of differential assignment to treatment due to unobserved factors. The U. Bound P-Value is calculated using the Rosenbaum Sensitivity Test for Wilcoxon Signed Rank P-Value, while the U. and L. Bound HL Est. are obtained using the Rosenbaum Sensitivity Test for Hodges-Lehmann Point Estimate.

References

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

Agatz, N., Campbell, A. M., Fleischmann, M., & Savels, M. (2008). Challenges and opportunities in attended home delivery. In *The vehicle routing problem: Latest advances and new challenges* (pp. 379–396). Springer.

Amazon (2019). About Amazon Fresh. https://www.amazon.com/gp/help/customer/display.html?nodeId=202052140. Accessed: 2019-04-26.

Autor, D. H. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of labor economics, 21*(1), 1–42.

Belavina, E., Girotra, K., & Kabra, A. (2016). Online Grocery Retail : Revenue Models and Environmental Impact Online Grocery Retail : Revenue Models and Environmental Impact. *Management Science*.

Bell, D., Gallino, S., & Moreno, A. (Forthcoming). Customer supercharging in experience-centric channels. Management Science,.

Boyer, K. K., Prud'homme, A. M., & Chung, W. (2009). The last mile challenge: evaluating the effects of customer density and delivery window patterns. *Journal of business logistics, 30*(1), 185–201.

Cachon, G., & Feldman, P. (2011). Pricing Services Subject to Congestion: Charge Per-Use Fees or Sell Subscriptions? *Manufacturing & Service Operations Management, 13*(2), 244–260. doi:10.1287/msom.1100.0315.

Cachon, G., Gallino, S., & Xu, J. (2018). Free shipping is not free: A data–driven model to design free-shipping threshold policies.. *Available at SSRN: https://ssrn.com/abstract=3250971*.

Cachon, G., & Swinney, R. (2011). The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. *Management Science, 57*(4), 778–795.

Calvo, E., Cui, R., & Wagner, L. (Conditionally Accepted). The value of disclosing product availability on retail platforms. Manufacturing & Service Operations Management,.

Campbell, A. M., & Savelsbergh, M. (2006). Incentive schemes for attended home delivery services. *Transportation Science, 40*(3), 327–341.

Chen, C., & Ngwe, D. (2018). Shipping fees and product assortment in online retail. *HBS Working Paper*.

Cochran, W. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, 295–313.

Gallino, S., Moreno, A., & Stamatopoulos, I. (2016). Channel integration, sales dispersion, and inventory management. *Management Science, 63*(9), 2813–2831.

Gaur, V., Fisher, M. L., & Raman, A. (2005). An econometric analysis of inventory turnover performance in retail services. *Management science, 51*(2), 181–194.

Gourville, J., & Soman, D. (2002). Pricing and the psychology of consumption.. *Harvard Business Review, 80*(9), 90–96.

Guo, L. (2010). Capturing consumption flexibility in assortment choice from scanner panel data. *Management Science, 56*(10), 1815–1832.

Harris, F. (1913). How many parts to make at once. *Factory,The Magazine of Management*, (10), 135–136.

Imai, K., King, G., & Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the Royal Statistical Society: Series A (statistics in society), 171*(2), 481–502.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review*, (71), 181.

Klein, R., Neugebauer, M., Ratkovitch, D., & Steinhardt, C. (2017). Differentiated time slot pricing under routing considerations in attended home delivery. *Transportation Science, 53*(1), 236–255.

Lantz, B., & Hjort, K. (2013). Real e-customer behavioural responses to free delivery and free returns. *Electronic Commerce Research, 13*(2), 183–198.

Lewis, M. (2006). The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing, 82*(1), 13–23.

Li, J., Granados, N., & Netessine, S. (2014). Are consumers strategic? structural estimation from the air-travel industry. *Management Science, 60*(9), 2114–2137.

Li, Y. P., Propert, K. J., & Rosenbaum, P. R. (2001). Balanced risk set matching. *Journal of the American Statistical Association, 96*(455), 870–882.

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

Rosenbaum, P. R. (2010). *Design of observational studies*: 10. Springer.

Sainathan, A. (2018). Customer differentiation with shipping as an ancillary service? free service, prioritization, and strategic delay. *Decision Sciences, 49*(4), 690–727.

Simonson, I. (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research, 27*(2), 150–162.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization, 1*(1), 39–60.

Yang, X., Strauss, A., Currie, C., & Eglese, R. (2014). Choice-based demand management and vehicle routing in e-fulfillment. *Transportation Science, 50*(2), 473–488.